THE DEVELOPMENT OF A MODEL FOR THE EVALUATION OF LOCAL UNIFIED SCIENCE PROGRAMS

By
THOMAS GADSDEN, JR.

A DISSERTATION PRESENTED TO THE GRADUATE COUNCIL OF
THE UNIVERSITY OF FLORIDA
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF EDUCATION

UNIVERSITY OF FLORIDA
1971

With my love

to my wife,

Sandy

and to my daughter,

Julianne

ACKNOWLEDGMENTS

The author would like to express his gratitude to the many people who have contributed to the successful completion of this study. A special thanks goes to his chairman, Dr. Luther Arnold, who has guided the author throughout his doctoral program and who has tirelessly provided the direction and careful, constructive criticism necessary for the undertaking of this dissertation. The author greatly appreciates the continual support and guidance he has received from his committee members, Dr. Vynce Hines and Dr. Richard Garrett. Thanks, also, to Dr. Daniel Swanson, who was a member of the author's committee until his retirement in 1971, and to Dr. N. E. Bingham who has been an inspiration and friend.

Special appreciation goes to the author's colleagues in the P. K. Yonge Laboratory School Science Department;-- to Paul Becht and Nancy Griffen for their patience and support, - to Virginia Allen who has put forth the tremendous extra effort necessary to develop, along with Dr. Donald Altieri and the author, the P. K. Yonge Integrated Science Program, - to Dr. Felicia West for her continuous encouragement and assistance, - and to Dr. Donald Altieri for his help in selecting the topic of this study and his continued interest and support in its development. Also, thanks to Dr. J. B. Hodges, Director of the P. K. Yonge Laboratory School, for his encouragement and understanding.

The author is deeply indebted to his typist and taskmaster, Christine Ward, for her excellent work, to Wendall Abbott for his

assistance in drawing the diagrams for this dissertation, and to Thomas Rymer and Gretchen Uzzell for their courageous efforts in proof-reading.

Deep gratitude and warm thanks go to the author's parents, Thomas and Doris Gadsden, for their ever present encouragement and moral support.

To his new daughter, Julianne, go thanks for her patience and cooperation and a promise that she will now have a chance to get to know her daddy.

And to his wife, Sandy, - more than these words can express. She has discussed and read and typed and corrected and encouraged. But more important, she has been the love and strength that is so essential. Without her it would not have been possible. To her, his love and appreciation, and the dedication of this dissertation.

TABLE OF CONTENTS

		Page
ACKNOWLEDGMENTS		iii
LIST OF TABLES		vi
LIST OF FIGURES		vii
ABSTRACT		viii
Chapter		
I.	INTRODUCTION	1
II.	JUSTIFICATION OF THE STUDY	6
III.	A GENERAL MODEL FOR THE PROCESS OF EVALUATION	20
IV.	EXPLICATION OF THE MODEL	31
V.	USE OF THE EVALUATION MODEL IN UNIFIED SCIENCE	121
VI.	A SYSTEMS MODEL FOR THE EVALUATION PROCESS	150
VII.	SUMMARY AND IMPLICATIONS	160
BIBLIOGRAPHY		166
BIOGRAPHICAL SKETCH		171

LIST OF TABLES

Table		Page
1.	Attendance at FUSE Conferences.	10
2.	Optimization of Evaluation Process.	65

LIST OF FIGURES

Figure		Page
1.	Flow Chart of Primary Components of Evaluation. . .	26
2.	Flow Chart of Evaluation Process with Feedback. . .	27
3.	Flow Chart for Identification of Criteria	33
4.	Flow Chart for Identification of Constraints. . . .	52
5.	Flow Chart for Optimization	60
6.	Stylized Flow Diagram of the Development-Evaluation Process.	68
7.	Flow Diagram of the Development-Evaluation Process for the Example Project	71
8.	Flow Chart for Design Development	73
9.	Flow Chart for Implementation	87
10.	PERT Network for Formative Evaluation of Laboratory Number Seven	94
11.	Flow Chart for Analysis	99
12.	Flow Chart for Reporting.	108
13.	Flow Chart for Decision-Making.	116
14.	General Systems Model	154
15.	Systems Model for Curriculum Development Evaluation.	157

Abstract of Dissertation Presented to the Graduate Council of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Doctor of Education

THE DEVELOPMENT OF A MODEL FOR THE EVALUATION OF LOCAL UNIFIED SCIENCE PROGRAMS

By

Thomas Gadsden, Jr.

August, 1971

Chairman: Dr. Luther A. Arnold
Major Department: Science Education

Evaluation should be an integral part of curriculum development. Too often, however, it is not. For small-scale, curriculum development projects having local concerns and unique purposes, evaluation is especially difficult.

In science, many curriculum development projects have been undertaken in the various disciplines on a national level. Now, there is a trend toward the development of interdisciplinary unified science programs that are designed primarily to meet the needs of a single school or single school system. These programs need careful evaluation. However they are limited by many serious practical constraints. Due to the uniqueness of unified science goals, there are few applicable instruments for data gathering. Comparison of interdisciplinary unified science programs with conventional disciplinary programs presents significant problems. The small-scale, local nature of these projects is frequently accompanied by insufficient funding, inadequate time, and inexperienced personnel not sufficiently skilled in evaluation. Thus, for small-scale, local unified science projects, a model or set of guidelines

is needed to assist in the evaluation process. A model or set of guidelines would provide an overall view of the process of evaluation, aid in the identification of criteria and constraints for evaluation, help in the selection of purposes and methods from the available alternatives, and assist in the identification and location of additional information about the evaluation process.

The purpose of this study was to develop a model to serve as a guideline in identifying evaluation procedures applicable to local unified science projects.

The development of the evaluation model began with the identification of the basic components of the curriculum development evaluation process and the examination of their interrelationships through flow charting techniques. Eight overlapping processes were identified as components of the curriculum development evaluation process.

I. Context Identification
   a. Identification of Criteria
   b. Identification of Constraints
II. Optimization
III. Design Development
IV. Implementation
V. Analysis
VI. Reporting
VII. Decision-Making

The identification of Criteria and Constraints occur simultaneously, otherwise, the process proceeds linearly through each component. A feedback is built in to allow for the assessment of changes in criteria or constraints.

The eight components were examined in detail in terms of the input, process, and output of each component. Flow charting was again used to illustrate the relationships among the various sub-parts of a component.

The application of the model to small-scale, local unified science projects was examined through the consideration of the unique criteria (needs, alternatives, and influences) and constraints of these projects and the effects of the criteria and constraints on various aspects of the process of evaluation.

In addition, a hypothetical example unified science project was followed through the entire evaluation process.

While flow charting techniques were found to be useful in examining the practical aspects of the curriculum development process, a general systems model was of more use in studying the process and its relationship to its environment on a theoretical level.

The study led to the conclusion that the process of curriculum development evaluation is a process of obtaining information for use in decision-making. The process can be characterized on a practical level through the use of flow charts and on a theoretical level through general systems theory. The resulting model can be used as a guide in developing and carrying through evaluations that are appropriate for the unique needs and characteristics of small-scale, local unified science projects.

CHAPTER I

INTRODUCTION

Unified science is an attempt to produce a more meaningful, more realistic, and more enduring science program by eliminating the traditional disciplinary barriers and studying natural phenomena through an interdisciplinary approach. The Federation for Unified Science Education defines unified science as a "...planned sequence of science experiences in which each unit of instruction utilizes subject matter from two or more traditional science disciplines" (Showalter, 1969). Each year increasing numbers of such unified science programs are being developed in small-scale, "local" projects involving no more than a single school or a single school system. This "grass roots" movement is rapidly producing a wide variety of unified science programs using various combinations of subject matter and processes from biology, physics, chemistry, the earth sciences, and in some cases, from areas such as anthropology, the social sciences, and psychology as well.

As small-scale, local unified science programs develop, there are many reasons for their careful evaluation. Among these reasons are the promotion of the effective and efficient use of the resources available to such curriculum development projects, the provision of feedback for the improvement of the developing programs, and the determination of the utility of the innovative approaches. However, the evaluation of these programs is complicated by the large number of alternatives that must be

considered in planning an evaluation and by the many constraints which limit the extent of any evaluation undertaken.

For example, the evaluator must select the purposes for his evaluation from among all of the possible purposes that an evaluation could have. Choices must also be made concerning the techniques of data collection, evaluation design, instrumentation, and the methods for the analysis and reporting of results. The constraints with which the evaluator must contend might include some serious limitations of budget and time. For the evaluation of unified science programs, few of the available instruments for data collection are appropriate. Many instruments that seem suitable from the title and description are found to be unsuitable upon close examination.

Basically, an evaluator has three alternatives in planning an evaluation: (1) to attempt to evaluate every identifiable purpose applicable to the project; (2) to attempt to limit the evaluation to only a few areas; or (3) to attempt to duplicate some evaluation carried out by a similar unified science project. The first choice is self-defeating for the small-scale, local project. Such an evaluation is a massive undertaking. If attempted with insufficient manpower, funds, time, or experience, the quality of all aspects of the evaluation is likely to suffer. The results of such an evaluation will tend to be misleading and of dubious value. The second alternative would be a good choice if guidelines were available to assist the inexperienced, local evaluator in making carefully considered decisions rather than arbitrary selections. Presently there is a crucial lack of such guidelines. The third possibility might produce a meaningful evaluation. But,it is unlikely that any two unified science projects would have exactly the same goals and approaches so that the same instruments and techniques

could be applied. In addition, the continued use of the same evaluation techniques, or techniques closely resembling them, assumes that no innovation is any different from previous approaches and that all innovations can be evaluated on the same basis. The desirability of the second of the above alternatives, and the importance of delimiting evaluations on a sound basis, indicate the critical need for better guidelines for making decisions concerning the evaluation of unified science programs.

Purpose

The purpose of this study is to develop a model and/or an instrument to serve as a guideline in identifying evaluation procedures applicable to local unified science projects.

Such a model or set of guidelines would have three purposes: first, to give project evaluators an overall look at the process of evaluation so that they could have a framework on which to build their evaluations; second, to present them with many of the possible alternatives and help them to make decisions about alternatives in light of their own unique situation; third, as these decisions are made, to assist the project evaluators in identifying and locating further information regarding the procedures they plan to use.

Schools or school systems considering the adoption or modification of an existing unified science program for their own unique situations would be likely to find this model or set of guidelines suitable for their purposes as well.

It could also be of use to the professional evaluator in that it could point out areas in need of new evaluation instruments or techniques, it could make more obvious some of the flaws in the evaluation process,

and could provide a framework within which much research on evaluation could take place.

Limitations

This study is intended to produce a model or set of guidelines suitable for use with small-scale, local unified science projects. While it may be of some use, it is not intended to be applicable to any other types of science programs or to a program of national scope. In addition, the study is built upon the definition of evaluation given in Chapter III.

Procedure

There are basically six parts to this study:

(1) The identification of the components of an evaluation model so that a better understanding of the overall process of evaluation might be obtained.

(2) The identification of criteria for an evaluation model for local unified science programs and the development of a means to assist evaluators in establishing criteria for their own evaluation programs. Examples of criteria for the evaluation model would be that the model presents an overall picture of curriculum evaluation with a maximum of clarity and understandability, and that the model permits a maximum of flexibility in the curriculum project during the evaluation.

(3) The identification of constraints and alternatives for evaluation. Constraints are the added factors that must be taken into account in determining the type of evaluation to be conducted and the factors which limit the range of permissible solutions. The amount of funds available and the duration of the project are two examples of such

constraints. The purpose for stating alternatives is to specify what possibilities remain after constraints are taken into account. For example, one alternative might be the use of a pre-test- post-test design. But, if the evaluation is not begun until near the end of the treatment period, then this alternative is essentially eliminated and the selection of evaluation design must be made from the remaining alternatives.

(4) The development of an instrument to assist in the optimization of the evaluation program. This instrument would assist the evaluator in selecting from the remaining alternatives those alternatives which most nearly meet the criteria of his evaluation program.

(5) The identification of sources of additional information needed by the evaluator. This would give assistance to the evaluator in making decisions that are beyond the scope of this study.

(6) The evaluation of the model or set of guidelines and its revision as necessary. This evaluation would determine the extent to which the finished model or set of guidelines meets the criteria established for it in Chapter II.

Summary

In summary, there is a growing movement in science teaching toward the development of local unified science programs. Along with this trend comes the need for careful evaluation. But, the local nature of these programs makes such evaluation difficult. Consequently, there is a need to develop a model or set of guidelines from the curriculum developer-evaluators of local unified science programs. This model or set of guidelines would give such persons an overall look at the process of evaluation, help them make decisions about alternatives in light of their own unique situations, and help them to locate sources of information for decisions that are beyond the scope of this study.

CHAPTER II

JUSTIFICATION OF THE STUDY

Rationale

This study is based upon two assumptions. The first assumption is that there is a need for the study due to a rapid spread of local unified science programs and due to the difficulties encountered by local programs in attempting to utilize present evaluation techniques. The second assumption is that it is feasible to produce an evaluation model or set of guidelines of this type. The following literature review provides support for these two assumptions.

History of Unified Science

The past decade and a half has witnessed dramatic changes in science teaching. Most obvious has been the development of the national project courses, the Physical Science Study Committee, Physics (PSSC), the Biological Science Curriculum Study, Biology (BSCS), the Chemical Education Material Study, Chemistry (CHEMS) and later the Earth Science Curriculum Project (ESCP), Intermediate Physical Science (IPS), and Harvard Project Physics (HPP), which now form the backbone of the science curriculum offered in most high schools in the United States. Less obvious but perhaps equally significant has been the development of an interdisciplinary approach to science teaching now called Unified Science.

The idea of a unified and continuous science program is not new. Such a program was advocated almost fifty years ago, but, with the

exception of a few scattered cases, it has not been achieved. (McCloskey, 1963, p. 3195). From 1936 to 1939 International Congresses for Unity of Science were held annually until World War II. Included in these conferences were such men as Niels Bohr, Percy Bridgman, Paul Langevin, Jean Perrin, and Bertrand Russell. In 1938, they began the International Encyclopedia of Unified Science with the copyright of the first chapter, "Encyclopedia of Unified Science" (Neurath, et al., 1938). By 1941, the trend toward a fusion of science courses had become observable and had resulted in the advent of physical science courses. These courses have steadily gained in popularity, but they usually consist of a combination of chemistry and physics offered only to students who do not plan to attend college (Smith, 1957, p. 350). The trend toward unification continued and by 1953 Paul DeH. Hurd was led to predict that high schools would soon be teaching Science I, II, III, and IV rather than departmentalized subjects, physics, chemistry, and biology (Slesnick, 1963). Instead, in the late 1950's, there was a reestablishment of departmentalization in science teaching, partly as a result of the new national science curriculum projects begun at that time. The teaching of unified science gained new life during the early 1960's as a number of relatively small-scale projects involving some degree of unification of the science disciplines developed independently and with very little federal support. Notable among the early projects are those of: Slesnick and Showalter, University School, Ohio State University; Leo Klopfer, University High School, University of Chicago; Michael Fiasca, Portland, Oregon; Morris Lerner, Barringer High School, Newark, New Jersey; Carl Pfeiffer, Monona Grove, Wisconsin; Leonard Blessing, Milburn, New Jersey; and Louis Bixby, St. Louis County Day School, St. Louis, Missouri.

During the 1960's, an increasing number of scientists and science educators recognized the need for a unified science offering in secondary schools. In regard to the national curriculum project courses, Leo Klopfer stated:

> ...each of the new courses falls short in presenting the whole of science as the intricately intertwined and intellectually integrated enterprise that it is. In fact, the new courses actually perpetuate and give additional support to the pattern of specialized science courses in the high schools (Klopfer, 1966, p. 27).

Fred W. Fox points out two other areas in which these new courses neither meet the needs of students nor those of society:

> (1) Designers of at least a few of the new secondary science programs have felt that the technology related to scientific progress is not worthy of inclusion in modern curricula even though technology is important to society, and (2) The problems of mankind have not warranted planned inclusion of these science courses, despite the fact that the scientific enterprise is related to both the creation and potential alleviation of the problems (Fox, 1966, p. 58).

James V. De Rose, at the time president of the National Science Teachers Association, stated in agreement with NSTA Curriculum Committee Report:

> ...it seems to me that it is necessary simply to say forcefully that a unified science program is certainly a desirable objective. To organize the science to be taught around a number of major conceptual schemes and process statements will give direction and concentrate the thinking of both the student and the teacher for a most efficient learning experience (De Rose, 1965, p. 83).

Paul Hurd has again stressed the importance of a unified approach to science teaching. "It appears that new courses will need to be invented that give more attention to the 'bridges' between the physical and biological sciences and humanities" (Hurd, 1967, p. 18).

The "first generation" of the contemporary unified science courses have now been in use nearly a decade and the unified science

movement is beginning to spread rapidly as second generation programs take shape. As one examines these programs, there is an amazing diversity evident in the approaches toward unified science. Each one seems to have been developed to meet the needs and potentials of the local area. This diversity is seen as an outstanding asset, and has led to many exciting exchanges of ideas. As long as diversity exists, unified science will remain dynamic, continually open to change, and responsive to the changes occurring in our society. The leaders of the unified science movement have expressed the hope that no one national unified science course will emerge, but that as courses develop they will be indigenous to single school systems or even to single schools, or will, at least, be readily modifiable to meet local needs (FUSE, 1969b). If the present state of the unified science movement is any indication, there is little doubt that, with a little encouragement, "local" unified science programs will begin to develop all over the United States.

Need for the Study

The development of unified science during the first half of the last decade was very slow indeed. As of May, 1968, only ten cases were reported in the current literature in which two or more courses had been combined to produce some sort of unified science program. At that time the author felt compelled to state that "there is no obvious trend toward fully integrated science teaching in the high schools" (Gadsden, 1968, p. 7). FUSE (Federation for Unified Science Education) which has a membership of about 400, in September, 1969, listed fifteen unified science programs currently in operation (FUSE, 1969a). In December, 1969, the New Jersey State Department of Education reported in the New Jersey Education Association Review that twenty-three schools in New Jersey

were involved in unified science curriculum projects (Showalter and Thompson, 1970, p. 2). Another indication of increasing interest in unified science is illustrated by a chart of attendance at FUSE annual conferences.

TABLE 1.--Attendance at FUSE Conferences

Time		Place	Attendance
August,	1966	Columbus, Ohio	8
August,	1967	Chicago, Illinois	16
December,	1968	Milburn, New Jersey	96
September,	1969	Monona Grove, Wisconsin	110
September,	1970	Portland, Oregon	140

(Showalter and Thompson, 1970, p. 2).

These reports, a recent literature search revealing seven additional unified science projects, and communication with the directors of several new projects show that there is now indeed a trend toward the teaching of unified science in secondary schools.

These unified science programs have shown a remarkable diversity in their bases for unification, in the grade ranges for which the programs are planned, in the specific objectives of the programs, and in the instructional methods used in the presentation of their approaches to unified science. The unification has in some cases consisted of the combining and restructuring of existing courses. In some instances the program has been built around interdiciplinary topics. In others the project has begun with a statement of interdiciplinary concepts which have then served as the foundation for the instructional program. However, in each case there is at least one common purpose; that is, to stress the interrelatedness and interdependence of the sciences and to move away from the segmentation of scientific knowledge and processes into compartments which have only limited functional value for the secondary student.

There has also been some diversity in the methods used to develop unified science programs. Some have been produced by task force teams whose primary responsibility was the development of such a program. Others have been designed by individuals or small groups given released time from other responsibilities to pursue a unified science project. Still others have been created by individuals in their "spare time." However, nearly all of these programs have been developed within and for a single school system or school.

A recent statement by Thomas D. Fontaine of the National Science Foundation indicates that there is a good possibility that even more of these local unified science programs will develop in the near future:

> Let me mention a few areas of curriculum development which we (NSF) believe should be considered at this time... 1. Interdiciplinary approaches to science curricula, e.g., multiple year sequences integrating natural and/or social sciences, or units of an interdiciplinary nature for use in existing courses (Fontaine, 1970, p. 210).

The desire of the FUSE membership to avoid a national unified science project and the apparent increase in the availability of funding for local, small-scale unified science programs could result in a number of new local unified science curriculum projects. As such projects develop, their careful evaluation is essential for the following reasons: (1) to determine the extent to which the resulting courses represent the goals of the project; (2) to determine the extent to which the materials produced are learnable; (3) to determine the extent of success or failure of this approach to curriculum development and science teaching; and (4) to determine the worth of continued effort in this direction.

Harold C. Hand takes this argument a step further. He stresses that unless three necessary conditions are met, "the likelihood is so

great as to approximate certainty that children will unwittingly be treated as pawns instead of sacred entities" (Hand, 1966, p. 4). The first of the three necessary conditions Dr. Hand advocates is that:

> ...the public must know instead of "wish-think" what the outcomes, for local children, of its innovative program practices are. This can be accomplished only by making rigorously controlled studies at the local level... (Hand, 1966, p. 4).

He continues:

> ...we are morally derelict unless we make carefully controlled factual studies at the local level of all curriculum innovations for local adoption or use (Hand, 1966, p. 6).

As local projects begin to undertake such evaluations they will probably be faced with the following alternatives: (1) a fairly large outlay of funds for professional evaluation; (2) an evaluation performed by a member of the curriculum development team for a lesser outlay of funds; (3) the permanent employment at the school system level of a person charged with evaluation of programs in that school system; or (4) no evaluation.

It has been shown that the fourth alternative is not desirable. The first alternative is apt to involve too much expense for the average local project. The second alternative would be a very likely choice. But, there are a number of serious difficulties that will be encountered with this alternative. It would be unusual for the person in a local project who is designated as the curriculum developer-evaluator to have a preparation in curriculum evaluation adequate to undertake an evaluation of this nature. It would also be unlikely that this person would be granted sufficient time to make the extensive search of evaluation literature necessary to prepare himself for such an evaluation.

Consequently, evaluations of this nature tend to either be based on the instruments that are most popular and most readily available at the time or they tend to attempt to include every possible aspect of the program, without first having a good understanding of the overall picture of such evaluation, the alternatives that are available, and the limiting factors that may be encountered.

The third alternative is promising, but it too is limited in that the local evaluation expert would be expected to conduct evaluations in all areas of curriculum development. Consequently, the curriculum developer would still have to bear a substantial responsibility for the evaluation. Unfortunately, as Guba points out, even "...the mere existance of an office or functionary within the schools charged with systematic evaluation is still rare" (Guba, 1969, p. 30). Guba goes on to offer additional evidence that evaluation today is "somewhat less than effective" at the local level.

> Local school districts rarely incorporate evaluation into any effort which they themselves fully control and finance. This is particularly evident when one consults proposed project budgets; if evaluation costs are included at all, they are contemplated only in very general terms, i.e., perhaps the salary of an evaluation "expert," or the cost of buying commercially available instruments (Guba, 1969, p. 30).

According to Guba, one of the more serious difficulties besetting evaluation is the "lack of meaningful and operational guidelines" (Guba, 1969, p. 30). If the curriculum developer is to play a major role in the evaluation of his own curriculum as the above argument suggests, then it is necessary that a model or set of guidelines for evaluation be available to him.

Although there were around forty experimental programs in unified science during the 1960's, only seven have reported any evaluation of

their curricula. Morris Lerner, in 1964, compared a combination of PSSC and CHEMS with a traditional program on the basis of subject matter attainment alone (Lerner, 1964, pp. 37-38). Louis Alcorta, while varying organizational patterns of teacher combinations and time, compared experimental unified curricula with traditional curricula on the basis of (1) ability to handle science problems; (2) science reading comprehension; (3) subject matter knowledge; and (4) perceptions of science and scientists (Alcorta, 1962, pp. 2030-2031). Two studies were undertaken concerning the unified science project at the Ohio State University School. Irwin Slesnick compared experimental and control groups on the basis of the individual's "rational image of the universe" (Slesnick, 1963, p. 302). Victor Showalter later did a follow-up study of students who had graduated from a unified science program concerning (1) interest in science; (2) scientific literacy; and (3) preparation for college science (Showalter, 1968, pp. 1-2). Michael Fiasca's study of the Portland Project was based on (1) critical thinking; (2) subject matter achievement; and (3) attitude change (Fiasca, 1967, p.2439). Carl Pfieffer at Monona Grove High School, Monona, Wisconsin, compared students who had studied science before the introduction of unified science with those who later took the unified science course. His data included the number of science courses taken by high school students and their scores on the Wisconsin Inventory of Science Processes (Pfeiffer, 1970, p. 4). Klopfer and McCann reported the only use of evaluative data for formative purposes. They compared the Natural Science program at the University of Chicago High School with the Time, Space, and Matter course using the Test on Understanding Science (TOUS) and a subject matter test. In addition to this comparison, they analyzed the results to identify

weak areas in the Natural Science course (Klopfer, 1969, p. 155). The Iowa Science and Culture Study is perhaps the most complete in evaluating the goals of the program. In that study, Cossman and Fitch compared experimental and control groups on the following criteria: (1) understanding scientific processes; (2) understanding scientists; (3) understanding science as related to society; (4) critical thinking ability; (5) substantive science knowledge; (6) importance of theoretical values; and (7) understanding of science as related to culture (Cossman, 1969, pp. 274-283).

At the 1970 Fuse Conference Victor Showalter, Michael Fiasca, Irwin Slesnick, Carl Pfeiffer, and the author discussed the possibility of developing an evaluation model for unified science programs. Concern was expressed that this model could have a restricting influence on the development of new unified science programs by imposing on them a particular set of values. However, with this caution in mind, it was decided that such a model could make a significant contribution to the evaluation of local unified science programs.

Feasibility of Study

The widespread development, during the past ten years, of national curriculum projects, especially in science, has produced a deepening concern for curriculum evaluation. As project directors attempted to revise the curriculum materials they were producing, they turned to evaluation in the form of feedback from teachers and students who were using preliminary versions of the materials. As funding for more projects became available through new federal assistance programs, such as Title I and Title III of the Elementary and Secondary Education Act of 1965, along with it came the requirement that educators evaluate

their new plans and programs. "As a consequence, many educators at all levels for the first time are having to cope with requirements for formal evaluation"(Showalter, 1968, p. 42). According to Stufflebeam, "The multitude of evaluation reports now available ... demonstrates that educators are expending significant amounts of time, effort, and money to evaluate their programs." Yet, "Many of the completed evaluation reports contain only impressionistic information"(Stufflebeam, 1969, p. 43). He attributes this to a lack of trained evaluators, a lack of appropriate evaluation instruments and procedures, and (most basic) a lack of adequate evaluation theory on conceptualizations pertaining to the nature of evaluation (Stufflebeam, 1969, p. 45).

However, during just the past few years there has been a marked increase in the efforts to improve curriculum evaluation. As a result of this effort,evaluation itself is becoming better defined and our general understanding of the process of curriculum evaluation has expanded to the extent that the study proposed is now feasible.

There are now a number of identified uses for curriculum evaluation. Michael Scriven suggests that evaluation is predominately either intrinsic (appraisal of the innovation itself) or pay-off (examination of the effects of the innovation on pupils) (Scriven, 1967, p. 54). Lindvall and Cox divide their evaluation into three segments on the basis of use: (1) individual pupil monitoring for the purpose of adapting instruction to individual needs; (2) formative, "...continuing evaluation of all elements of a developing educational program as an aid to the development process" (Lindvall and Cox, 1970, p. 2); and (3) summative, "...the evaluation of the results produced by an educational program for the purpose of making judgements concerning its value"

(Lindvall and Cox, 1970, p. 2). Several curriculum evaluators have suggested that evaluation consists of a number of distinct phases or steps. Garlie Forehand identified five such phases: (1) identifying goals; (2) formulation of decision rules; (3) construction of measuring instruments; (4) collection and analysis of data; and (5) making the decision (Forehand, 1970, pp. 29-30). Edward Brown similarly lists five phases of evaluation: (1) tentative procedural and behavioral objectives; (2) monitoring the program; (3) revising the program; (4) program assessment; and (5) program implementation (Brown, 1970, pp. 68-69). Four strategies for evaluation are given by Daniel Stufflebeam: (1) context evaluation; (2) input evaluation; (3) process evaluation; and (4) product evaluation (Stufflebeam, 1968, pp. 62-65). Similar phases have been identified by Robert Randall (Randall, 1969, p. 41) and by Marvin Alkin (Alkin, 1970, p. 5).

There are also now several data collection designs that have been described adequately enough to be of use in curriculum evaluation. Campbell and Stanley have carefully explained the various types and uses of experimental designs and quasi-experimental designs (Campbell and Stanley, 1963). Light and Smith have explained more fully the *post hoc* survey designs and they have also introduced an alternative to the *post hoc* experimental design, called the exploratory-experimental design, and an alternative to the *post hoc* survey that they call the sequential survey (Light and Smith, 1970).

Several outstanding books and articles have been written concerning the overall picture of curriculum evaluation. Hulda Grobman's book, *Evaluation Activities of Curriculum Projects*, does an excellent job of presenting many of the limitations and alternatives that could develop

in a curriculum project evaluation (Grobman, 1968). An article by Garlie Forehand presents a clear overall picture of the role of evaluation in decision-making (Forehand, 1970). The American Educational Research Association (AERA) Monograph by Lindvall and Cox gives a complete look at the evaluation of a single project, Individually Prescribed Instruction (IPI), illustrating the formative, summative, and individual progress uses of evaluation (Lindvall and Cox, 1970).

Along with the advances in curriculum evaluation indicated above have come a number of models and paradigms for evaluation. Some of these are: a generally accepted statement of a national curriculum project model (Jacobson 1970, p. 223); Crittendon's General Summative Model (Westberry, 1970, p. 251); Stufflebeam's Context, Input, Process, Product (CIPP) Evaluation Model (Stufflebeam, 1969, p. 60); Brown's Process Evaluation Paradigm (Brown, 1970, p. 69); Stake's Evaluation Matrix (Stake, 1969, p. 16); and the Conceptually Oriented Program in Elementary Science (COPES) Evaluation Model developed by Shamos and Barnard (Shamos and Barnard, 1970, p.3). Each of these models offers a slightly different look at the evaluation process and each one will provide needed guidance for the curriculum project evaluator. None of these models, however, is aimed toward the evaluation of <u>local</u> curriculum projects nor toward the evaluation of unified science programs. Yet, without these models and the other advances in evaluation listed above the development of an evaluation model for local unified science projects would be an immensely difficult if not impossible task.

Summary

Hulda Grobman makes perhaps the strongest statement in favor of this type of undertaking in <u>Developmental Curriculum Projects: Decision Points and Processes</u>:

The temptation of the critic of project evaluation is to suggest the grandiose--to ask for a "complete" evaluation. The accomplishment of a complete evaluation is an impossibility. It would take too long. It would be too expensive. The materials being evaluated would be obsolete before such an evaluation could be completed. And there are insufficient trained staff to perform evaluation tasks for all developmental projects currently in existance.

It would appear more realistic to ask for more complete evaluations, for evaluations that are better geared to serve the purposes and needs of the individual projects, and for the development of more evaluation systems--for guidelines and structures that can be applied, with or without modification, to various aspects of curriculum evaluation. Despite the danger that an inappropriate system will be foisted on a project, or that unimaginative approaches to evaluation problems will be designed, such systems can provide starting points, and as a number of parallel systems are developed, projects can be more easily aware of the evaluation choices they may make and the ramifications of these choices (Grobman, 1970, p. 224).

CHAPTER III

A GENERAL MODEL FOR THE PROCESS OF EVALUATION

The first step toward understanding a process as complex as curriculum evaluation is to obtain a clear overall view of the whole. That is, the identification of the primary components of the process and the description of their relationships to one another. Once this overall view is obtained, each part of the whole can be examined without losing sight of the importance of any single element and the role it plays in the total system and without emphasizing any one element to the exclusion of others.

In order to delimit the process under consideration, it was necessary to obtain a definition of evaluation that would be inclusive of any aspect of the evaluation of local unified science curriculum development projects. Much confusion about evaluation has resulted from a lack of discrimination between research and evaluation. While both of these are seen as part of a continuum, it is useful for the purposes of this study to make a distinction between the two. Consequently, research is identified as a basic study of the relationships between variables in which the variables are to some extent isolated and controlled. Usually such research is motivated and guided by some hypotheses held by the researcher. The scope, timetable, and instruments used are those appropriate to the investigation of the hypotheses.

On the other hand, evaluation is more akin to the applied study of some treatment carried out in an otherwise uncontrolled or little controlled setting and for which the treatment itself may be changing. In the case of the evaluation of developmental curricula "the convenience and requirements of the project set the constraints on the evaluation" (Grobman, 1970, p. 174).

Definitions of evaluation seem to fall into five categories:

(1) Measurement--Many definitions, both implied and extant, equate evaluation and measurement. According to these definitions, testing and other forms of measurement are identical with evaluation.

(2) Professional judgment of worth by a professional. (e.g., "My evaluation of this student's progress is 'B'," or "Our evaluation of this condition is that it needs improvement in two areas.")

(3) Congruence--Much evaluation presently conducted attempts to determine the congruence between student performance and previously determined objectives.

(4) Feedback--Evaluation is defined as the gathering of data which areto be used for the modification of programs in use and under development.

(5) Decision-making--In much of the current literature evaluation is being defined as a part of the decision-making process.

Each of the first four categories concerns only a part of the evaluation process. None is sufficiently comprehensive to include all of the roles that evaluation is now playing. Consequently, for the purpose of this study a definition developed by the Center for the Study of Evaluation, University of California, Los Angeles, California, and which falls into the fifth category will be used. "Evaluation is the process

of ascertaining the decision areas of concern, selecting appropriate information, and collecting and analyzing information in order to report summary data useful to the decision-makers in selecting among alternatives" (Alkin, 1970, p. 2). Decision-makers are those people who will in some way make decisions affecting the project or its future. Since the decision-makers may be internal (project staff members) as well as external to the project and since the total evaluation program may be concerned with several different decision areas, this definition is inclusive of each of the other four categories mentioned above. Yet it is specific enough to be of great use in helping to understand the evaluative process.

This definition provides the basis for the next step, identification of the primary components of the evaluation process and the interrelationships among those components.

Evaluation consists of eight overlapping processes:

I. Context Identification

a. Identification of Criteria

b. Identification of Constraints

II. Optimization

III. Design Development

IV. Implementation

V. Analysis

VI. Reporting

VII. Decision-Making

I.a. Identification of Criteria refers to the process of identifying all of those purposes toward which some part of the evaluation could be justifiably aimed under ideal conditions. The "potential purposes of the project evaluation" are selected from the set of all

possible "purposes for evaluation." The goals (or purposes) of the project are just one part of the criteria on which this choice is based. The process is intended to answer the question,"If evaluation of this project could be conducted without limitations, what questions should be considered?" In addition to identification, this process would also assign a priority ranking to each potential purpose.

I.b. Identification of Constraints refers to the process of identifying and determining the relative importance of factors which limit the range of possible approaches and solutions to the decision problem.

II. Optimization refers to the process of comparing the criteria that have been established with the constraints identified in order to develop the most effective and most functional set of purposes for the evaluation of a particular project. These are the final "purposes of the project evaluation." This also implies determination of the optimum relationships among these purposes of the project evaluation in terms of how they might affect one another.

III. Design Development refers to the process of determining how each of the purposes of the project evaluation will be met. This includes specification of the methods, instruments, and scheduling to be used in the data collection and a general consideration of the methods, instruments, and scheduling anticipated for the analysis of data and reporting of conclusions. When necessary this component would also include the development of new instruments and techniques.

IV. Implementation refers to the process of creating and gathering data and organizing these data into some orderly and synthesizing form. This also implies the performance of any logistical steps necessary to facilitate the creation and gathering of data.

V. Analysis refers to the process of drawing meaning from the collection of raw data. The process is roughly divisible into three steps: (1) organization of raw data to facilitate analytical procedures; (2) analytical processing of data; and (3) when applicable, determination of the extent of significance (both statistical and practical) of the resulting conclusions.

Analytical processing implies a number of different kinds of operations. Among these would be: (1) comparison of the data with norms, pre-established goals or criteria, values or expectations; (2) identification of the elements of a phemonenon and establishment of a hierarchy among those elements; (3) determination of relationships among variables; and (4) recognition of organizing principles basic to the subject of the study.

VI. Reporting refers to the process of interpreting conclusions and translating them into information that can be meaningfully communicated to the intended decision-making audiences. Different reports should be developed for each different decision-making audience. These communications could be designed for different levels of sophistication and could utilize different media forms depending on the needs and background of the decision-maker involved.

The process might also include the identification of important implications regarding the project that the evaluator feels could be justifiably drawn from the conclusions of the evaluation.

VII. Decision-making refers to the process of the decision-maker in considering the evaluative information with respect to that area of the project with which he is concerned. Then, together with any information regarding this area obtained from other sources, and the internal and

external pressures acting upon the decision-maker, he determines what action should be taken regarding the area concerned.

There is extensive overlap among these eight component processes. Basically, however, they are related sequentially both in time and function. The Identification of Criteria and the Identification of Constraints, components I.a. and I.b., are exceptions in that they are more likely to occur simultaneously. See Figure 1.

Each of the components in Figure 1 consists of three sub-parts: input, process, and output. An input is a factor that will in some way influence the process. An output is the product resulting from the process. The goal of the process in each case is to produce the desired output in the context of the inputs of the situation. In a sequential relationship such as this, the output of each component becomes part of the input of the next component. Thus, each component is dependent upon the adequate functioning of the previous component. A more detailed description of the input, process, and output will be developed in the next chapter.

One of the major distinctions of curriculum project evaluation is that the developing curriculum is continually changing, partly as a result of the evaluation itself. Thus, the object of the evaluation study is not a static entity. It is dynamic. Consequently, the evaluation must also be dynamic. As decisions are made and the curriculum changes, both the criteria and constraints may change. Some provision must be made in any model of the evaluation process to encourage such flexibility.

The feedback system shown in Figure 2 is introduced to illustrate how this flexibility might be accomplished. Between each component

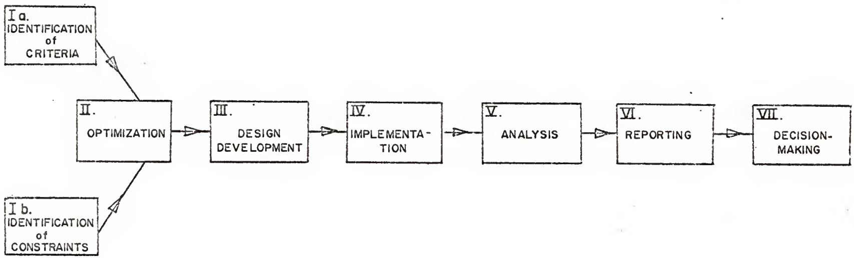
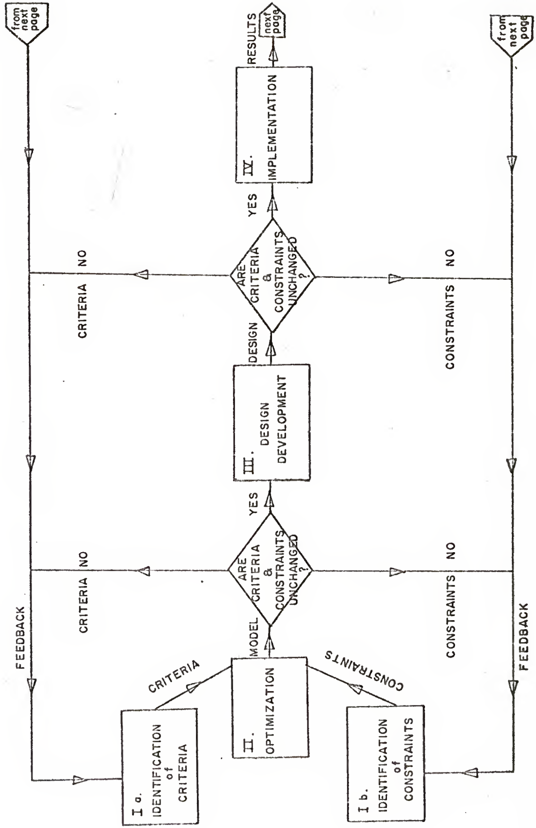


Figure 1.--Flow Chart of Primary Components of Evaluation



I a. IDENTIFICATION of CRITERIA
II. OPTIMIZATION
I b. IDENTIFICATION of CONSTRAINTS
CRITERIA
CONSTRAINTS
MODEL
ARE CRITERIA & CONSTRAINTS UNCHANGED ?
YES
III. DESIGN DEVELOPMENT
DESIGN
ARE CRITERIA & CONSTRAINTS UNCHANGED ?
YES
IV. IMPLEMENTATION
RESULTS
next page
CRITERIA NO
CRITERIA NO
CONSTRAINTS NO
CONSTRAINTS NO
FEEDBACK
FEEDBACK
from next page
from next page

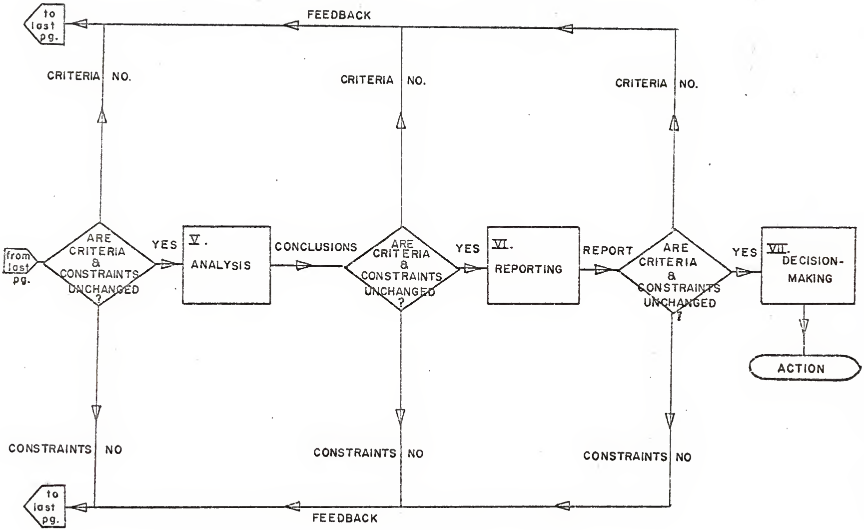


Figure 2.--Flow Chart of Evaluation Process with Feedback

process, both criteria and constraints are reexamined. ("Are criteria and constraints unchanged?") If they have not changed since they were last considered, then the evaluation can proceed. However, if either has changed, then the evaluator must consider the effects of the altered criteria and/or constraints on the evaluation and their effects on the steps he has taken up to that point.

Not every change would be of such significance as to warrant major changes in the evaluation. Certainly to formally go through the entire process every time some minute change occurs would mean a never-ending process that would never produce results. But, this feedback process does call for continuous monitoring of both criteria and constraints, so that as important changes do occur, the evaluation can be quickly responsive to those changes. Otherwise the evaluator may persist in obtaining data that are no longer desired or in aiming a report toward an audience that is no longer in a decision-making position. Such oversights would, at best, be wasteful and unnecessary. At worst, they may be extremely detrimental or even disastrous to the project.

The evaluator must remember that evaluation is a service function for the project and for the decision-makers concerned with the project. As such, it must in no way limit the innovative development it seeks to serve.

Summary

This chapter has considered a definition of evaluation suitable for the new roles of evaluation in developmental curriculum projects and has identified the major components of the process of evaluation consistent with this definition. The series of processes represented in Figure 2 is not specific to any one type of evaluation. All types of

evaluation (context, process, formative, summative, etc.) follow this same general procedure whether the evaluation lasts a single day or several years, whether the decision-maker is a member of the project staff or a person external to the project who in some way controls the project's future, and whether the evaluation is formal or informal. In limited cases, some aspects of the evaluation may be carried out entirely in the mind of the evaluator, while in other more complex cases evaluation may require explicit, formal attention to each component of the process.

There is considerable overlap among these component procedures, in function as well as in time. For example, one part of the total evaluation may be in the process of analysis while another part of the evaluation is just entering the implementation process. Reports and decisions regarding one phase may have already been completed while the development of instrumentation for another phase is underway. The identification of criteria and constraints probably occurs simultaneously to a great extent. It is also most likely that the optimization process begins long before the criteria and constraints have been fully established. However, this general model does provide an overall view and a fair picture of the sequence of processes that take place in any evaluation.

The representations of the evaluation process in this chapter have been in the form of flow charts. This format is useful in describing processes since it is clear-cut and can be easily understood. But, in order to present a more complete picture of the evaluation process and its relationship to the environment in which it takes place, it is more useful to illustrate the process as a systems model. The development of such a model will be examined in Chapter VI.

CHAPTER IV

EXPLICATION OF THE MODEL

Once the framework has been set and a comprehensive picture of the overall process of evaluation has been developed, then the parts that make up the process can be meaningfully examined in detail. In this chapter, the primary components of the process of curriculum development evaluation will be examined by describing the input, process, and output of each component. Flow chart diagrams accompany each explanation in order to provide quick reference and to assist the evaluator in obtaining a fuller understanding of the substructure of each component.

The diagrams consist of several sources of input ( ) which contribute information to the process, a brief general description of the process ( ), and an output from the process ( ). Arrows indicate the direction of flow ( ). At the top and bottom of many of the charts are continuation blocks ( ) indicating that the process continues into the next primary component. The output from each component, except the last, serves as input to the next component and is also shown as an input on the diagram from that component ( ). The Identification of Criteria (I.a.) and the Identification of Constraints (I.b.) occur simultaneously. Consequently, they both continue directly to Optimization (II.).

I.a. Identification of Criteria

Input

In Figure 3, the input for identifying criteria for the evaluation has been grouped into three categories; (1) Needs, (2) Alternatives, and (3) Influences.

The needs of the curriculum project evaluation, according to the definition of evaluation discussed earlier, are decision-making needs. According to this definition, the role of the evaluation is to obtain and provide decision-makers with the information needed to make wise decisions concerning the curriculum project. This input is not necessarily readily available to the evaluator. Instead, it is more likely that he will have to make a conscious effort to obtain input regarding needs. The identification of these needs implies three interrelated steps:

(1) a determination of the decision questions that the project will immediately or ultimately face;

(2) the identification of the persons and institutions to make these decisions (the decision-makers);

(3) a determination of the information requirements of the decision-makers.

To aid the evaluator in more fully understanding the nature of this input, a number of different types of decisions have been identified that may be involved in a curriculum project. Later in this chapter, decision-makers usually associated with each decision type will be identified.

A. Developmental decisions are directly concerned with the innovation itself, with respect to the adequacy, appropriateness, feasibility, correctness, sequencing, and/or optimization of the program

I a. IDENTIFICATION OF CRITERIA

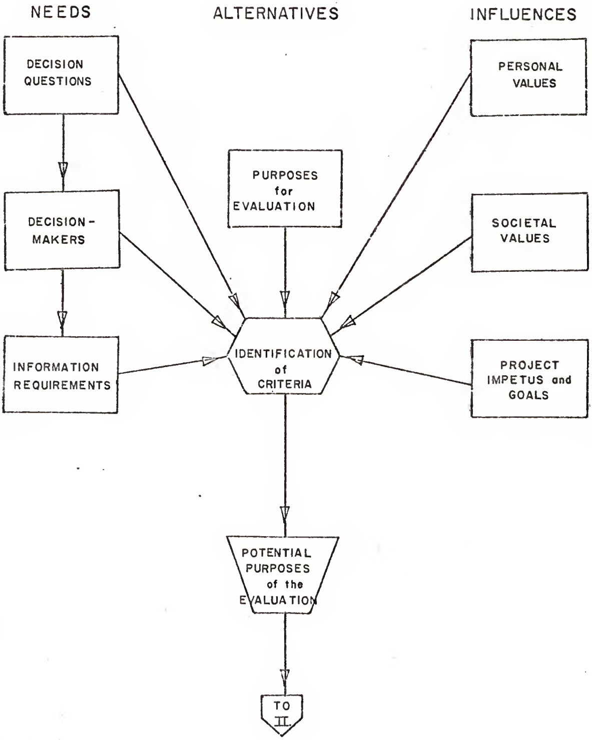


Figure 3.--Flow Chart for Identification of Criteria

or its parts. The decision is one of approval (accept, reject, modify).

B. Operational decisions concern the project procedures and schedule.

C. Administrative decisions concern the effects of the innovative curriculum and the project on the school in which they operate. They are especially concerned with the project's effects on other courses or other functions of the school, the cost of the project to the school (all costs, monetary and otherwise), and how the innovation fits with the purposes of the school. In addition, this type of decision would consider the extent to which the project meets commitments it has made in agreement with the administrative decision-makers. The decision here is one of permission (go, no go, modify program).

D. Policy decisions are those concerned with the extent to which the innovative curriculum and the project fit with the societal purposes and goals, with how they relate to parental goals for the children involved, and with possible detrimental effects that the program might have.

E. Support decisions include those related to appropriation and allocation of funds, space, and time, and those related to public opinion.

F. Professional decisions are those concerned with how the products of the project and curriculum correspond with the goals of the profession (science education), with the processes used in the development of the curriculum innovation, with the adequacy of the evaluation, with the success of the innovation in meeting its own goals, and with the implications of the project and evaluation for other areas of science education.

G. Dissemination decisions are concerned with the saleability of the program to others, the effectiveness of the diffusion process, the form in which the new program will be most suitable to other schools.

H. Adoption decisions concern the applicability of this innovation to other school situations, the success of the innovation in meeting its own goals, the program's success and cost in comparison with other possible approaches (especially any it might replace), the extent and ease with which the program fits with the goals of interested schools, and the ease with which the program might be adapted to the needs of the interested schools.

Nearly anyone connected directly or indirectly with a curriculum development project could be a potential decision-maker. This knowledge, however, is of little help in determining the criteria for the evaluation. A list of such decision-makers may be of help:

1. Project Personnel (Includes director, staff, evaluators, and in some cases the cooperating teachers.)
2. Professional Educators
3. Other Science Teachers
4. Students
5. Parents
6. Lay Public
7. School Faculty
8. Principal
9. District Supervisor
10. District Superintendent
11. School Board
12. State School Officials
13. Other School or School District Administrators
14. Funding Agency
15. Publishers
16. Political Power Structures (formal and informal)

The main usefulness of this list is to insure that no important decision-maker is overlooked in establishing the criteria. Some on this list may seem, at first, to play a trivial role as a decision-maker. Each can, however, in the appropriate circumstances, be a very

powerful decision-maker. For example, students have very direct evidence about the success of the program, but their basis for comparison might be goals other than the goals of the project. To neglect their roles as decision-makers could, in some situations, lead to serious difficulties and misinformation in other aspects of the evaluation.

More useful, perhaps, would be a consideration of each type of decision along with the decision-maker who will most likely make the decision.

Developmental decisions are most often made by project personnel. Operational decisions, too, are most often part of the function of project personnel. Administrative decisions are made by school district supervisors and administrators, the school principal and other school administrators, and in some cases the school faculty. Policy decisions are most likely the function of the school board, the district superintendent, and other local, state, and federal legal bodies. Parents and members of informal political power structures may also play significant roles in policy decisions. If important decisions are to be made, it would be worthwhile to determine which decision-makers truly wield the power concerning a curriculum development project.

Support decisions should be considered with respect to the types of support. Appropriation support decisions might be made by public or private funding agencies and by various legal bodies. Allocational support decisions are more likely to be made by school and district administrators. In each of these two cases there also may be a less formal decision-making structure that is no less important. Students, parents, other teachers, and the lay public (especially mass media personnel) are the decision-makers for public opinion support decisions.

Professional decisions are made by science educators, members of professional societies, and journal editors. Dissemination decisions are made by publishers and funding agencies as well as by project personnel. Adoption decisions are primarily the concern of administrators of other schools and school districts.

It may be difficult to determine the specific decision needs of the various decision-makers. In some instances, an evaluator will have to make an educated guess about these needs. In some cases a personal interview or questionnaire might help. Often getting to know the decision-maker and becoming able to communicate freely with him will be of great importance. Once communication lines have been established, frequent interaction with the decision-maker could provide insight that would enable the evaluator to recognize changes in information needs and to have a better basis for establishing priorities. For example, in some cases the decision-maker may already have sufficient information. To provide more would be wasteful of limited resources. In another case, the evaluator may notice a need for information that the decision-maker is hesitant to request. These are seldom acknowledged but very important problems that the evaluator must consider.

In the past few years, a substantial number of alternative purposes for evaluation have been identified. In fact, many of the recent articles in evaluation literature are treatises on one or more purposes for evaluation that have, in the past, been given inadequate consideration or even been ignored. One useful way to consider these alternatives is in relation to time. That is, at what point in time relative to the progress of the project is this information needed by the decision-maker concerned? A fairly complete set of alternative

purposes may be obtained by examining articles of the type mentioned above and grouping alternative purposes with respect to time as illustrated below.

Before the curriculum development and trial use of an innovation formally begins, decision-makers may require evaluation conducted for the following purposes:

1. Assessment of Social Setting. (Alkin, 1970.) This includes consideration of the social and political atmosphere, identification of community goals, general education goals, school goals, and educational needs; and determination of the relationship of each of these to the project goals and plans.

2. Context Assessment. (Stufflebeam, 1969.) This includes the identification of the specific problems with which the project will be concerned, and the establishment of project goals and overall plans.

3. Antecedent Assessment. (Stake, 1969.) The purpose of antecedent assessment is to establish bench marks for future reference and description. Some of the many items that could be considered are: description of the learner; description of the teacher; description of the school setting; adequacy of the classroom and equipment; class size; sex of students and teacher; class period length; class norms; individual ability and achievement levels; and previous exposure of students and teachers to the subject area and related areas.

During the actual curriculum development and trial phase of the project, decision-makers require information from evaluations designed for a number of different purposes. For example:

1. Assessment of the Developmental Process. (Grobman, 1968.) Such evaluation could be concerned with two aspects of the developmental

process, the process followed by the project in creating the innovation and the process followed in the diffusion of the program developed. The monetary costs, allocation of personnel, time expended, and effectiveness of the process would be of interest in each case.

2. Study of Transactions. (Stake, 1969.) This refers to the determination of what is actually occurring in the classroom and the relationship between these observations and the intentions of the project.

3. Formative. (Scriven, 1967.) The purpose of formative evaluation is to provide information for the improvement of the program being developed while the development is under way. Grobman suggests that such evaluation could be concerned with the micro-level (specific activities, approaches, sequencing, etc.) or the macro-level (the entire program or major components of the program) and would need to consider each of the three domains (cognitive, affective, and psychomotor). (Grobman, 1970.)

4. Individual Pupil Monitoring. (Lindvall and Cox, 1970.) This purpose becomes important when the project is concerned with the effects of the program on individual students rather than, or in addition to, its effects on groups of students or on the average effects on all students.

5. Identification of Dysfunctional Outcomes. (Grobman, 1970.) Monitoring for detrimental effects is seldom included in curriculum project evaluation. Yet, as Grobman points out, "they may be far more detrimental than is generally realized ... dysfunctional outcomes may outweigh any gains that are achieved" (Grobman, 1970, p. 200).

Many decisions are made after a project has completed a trial run or after the entire project has been completed. Such decisions

may require the use of data that have been collected throughout the duration of the project. Evaluation aimed to assist in such decisions may be grouped according to the following purposes.

1. Summative. (Scriven, 1967.) Summative evaluation includes any evaluation whose purpose it is to describe the finished output of the project. Some evaluations that could be included in this category are:

a. product and activity description, so that the product can be replicated to the extent possible and desired by adopting schools or by other projects;

b. extent of goal achievement, for the purpose of program certification in each of the three domains;

c. comparative effectiveness, again for program certification;

d. determination of causality, that is the relationship of results to the operation of the program, to improve chances of repeatability; and

e. identification of unintended outcomes both advantageous and detrimental.

2. Long-Term Effects. These effects can be studied in all three domains relative to the effects of the program on students and the effects of the project on science education.

Criteria can be influenced intentionally and unintentionally, desirably and undesirably by a number of factors. Whether or not these factors can or should be controlled, they should be acknowledged. Criteria should be selected with some idea in mind of how that selection is affected by such influences. Some of the more important of these

factors are personal values, societal values, and the impetus and goals of the project.

The personal values of the project director and staff and of the evaluator will have significant effects on the selection of criteria and on the entire evaluation as well. In research, the effects of personal values on experimentation are called "biases" and are treated as potential sources of error. Since evaluation deals with "real," uncontrolled situations, such biases may be an integral part of the situation and probably play an important role in the development of the program itself. Some recommend that evaluation be carried out by evaluators external to the project for just that reason. The assumption is that the value system of the external evaluator will not be as intimately tied to the development of the program as would be those of the project personnel. Consequently, he would be able to evaluate the program more objectively. In the situation that the project director or the evaluator is also the decision-maker, then his values must be considered as legitimate factors in establishing the criteria. When he is not the decision-maker, the influence of his value systems may mislead the evaluation efforts to produce information that is not desirable or acceptable to the decision-maker. It may even lead to a poor choice of criteria at the outset of evaluation. For example, when attempting to establish criteria, personal values might effect the evaluator's perception of the decision-maker's needs, might effect the priority given the various needs identified, or might cause the evaluator to automatically eliminate potential purposes for the evaluation from consideration.

To attempt any comprehensive listing of personal values is beyond the scope of this paper. However, a number of examples are given

below in the hope that they will provide some insight into how the reader's own values might effect his evaluation activities. Some possible personal values or things valued would be: cooperation, competition; knowing, wondering; change, stability; newness, tradition; quality, quantity; individuality, socialization; independence, conformity; ends, means; technological production, environmental conservation; regimentation, flexibility.

An evaluator needs to understand his own values with respect to the project and with respect to the values of the project personnel. If he is internal to the project (that is, a member of the project staff), then such understanding will permit greater objectivity. If he is external to the project (hired for the purpose of evaluation), then this will allow him to better anticipate and avoid any misunderstandings of the project goals and any points of conflict between his value system and those of the project personnel.

Other sources of influence in the selection of criteria for the evaluation are the societal values and character of the community in which the evaluation is taking place. Presently, there is substantial conflict and confusion concerning values in our culture. There is little acceptance of common values and frequently little agreement between our espoused values and our characteristic actions. In nearly every American community a number of different value systems can be identified and associated with various socio-economic groups, political groups, interest groups, business groups, and religious groups. Each different group may represent a different value system so that there is no one set of accepted community values. However, there may be one or more systems of values that can be identified as predominant in the community.

Whether or not their values are predominant, various groups may try to influence the selection of criteria. Pressure groups or political bodies may intentionally attempt to direct the evaluation toward some particular aspect of the program. For example, an interest group might insist on knowing the extent to which the project is teaching values with which the group agrees. Various decision-makers who feel that they represent the community may desire that their decision needs be given a high priority.

The social climate of a community may be evidence of harmony or conflict among the values of the various groups mentioned above. There will be some communities in which traditional values are evident. In such communities, the way in which the participants in an evaluation respond to the evaluator's attempts to obtain information will be influenced by their traditional values and by the extent to which the aims of the evaluator correspond with those values. In a community in which value conflicts are prevalent, it may be profitable to the project both educationally and politically for the evaluator to consider the influence of the project on the values of the participants. For example, there is currently much controversy about environmental problems and their solutions. In communities deeply involved in such controversy a higher priority might be assigned to the evaluation of the program related to the environmental controversy.

In another respect, the social climate or various pressure groups can effect the selection of evaluation criteria by effecting the project itself. For example, in a number of communities the teaching of sex education in public schools has been strongly challenged. In another case a community banned the use of the course, Man: A Course of

Study, in its school system. In at least one state, many of the national curriculum project courses were banned because of their indirect association with the Biological Sciences Curriculum Study which teaches evolution. Any other project developing materials for use in these communities which are similar in nature to the programs mentioned, or which might be seen as similarly threatening, could face similar difficulties. Such pressures could result in distorted and misleading evaluative data. Evaluation criteria in such a situation must be carefully chosen.

Many of the influences of societal values presently may not be evident, especially since curriculum development projects often proceed quietly and in partial isolation. As the demand for accountability in education increases and becomes more a part of curriculum development, so will the obviousness of societal influences.

To adequately select criteria, the evaluator must understand why a project exists and what it hopes to accomplish. This assumes that the project personnel have carefully documented the impetus for the project and have clearly established the goals which it hopes to achieve. Certainly, as new impetus arises, goals may shift. As they do, it is the responsibility of the evaluator to remain aware of these shifts and to adjust the criteria for the evaluation in order to be consistent with the new foci.

The impetus for curriculum development projects usually comes from frustration with the present curriculum and from an awareness of the existence of social forces that are significantly affecting our lives, and yet, are not adequately treated in the current curriculum. Some of the social forces that are now of primary importance to our future are: rapidly changing values and morality (especially concerning

sex and drugs); the changing status of minorities; social recognition of environmental degradation; student dissent; student alienation; urbanization; rapid scientific advancement; the knowledge explosion; women's liberation; the possibility of world destruction; cybernation and computerization; and the rapid change in nearly all aspects of life.

From such social forces curriculum projects have developed a wide range of goals. A more careful analysis of the impetus and goals of local unified science projects will be considered in the next chapter.

Process

The process of the identification of criteria consists of the consideration of the alternatives in terms of the evaluation needs and the internal and external influences on evaluation. This entails finding answers to the following questions:

(1) What decisions are appropriate to the project or are necessary for its continued operation and improvement?

(2) Who are the decision-makers for these decisions?

(3) What are the informational needs of these decision-makers?

(4) What alternative purposes for evaluation are available?

(5) Which of these alternative purposes are suited to meet the information needs?

(6) Which alternatives are not appropriate in relation to the project goals and the community needs?

As potential purposes for the evaluation are identified, the following question must also be answered. What priorities should be placed on these possible purposes for evaluation?

Throughout this process the evaluator must continually be aware of the biases introduced by the personal values of both the project director and himself.

Output

The result of the process of the identification of criteria should be the specification of the potential purposes for the evaluation that are deemed appropriate and needed. This should include specification of the type of information, the decision-maker, the focus, and the specific questions involved. In addition, the priority assigned to each potential purpose and the rationale for the assignment of that priority should also be indicated. A single, hypothetical, curriculum development project evaluation will be used as an example to illustrate the output of each of the component processes considered in this chapter. The hypothetical project is concerned with the development of interdisciplinary science laboratory experiences and materials for use in the unified science program of the sponsoring school, and possibly for use in other schools having unified science programs. The discussion of this example project is not intended to suggest that the evaluation indicated for this project represents the ideal evaluation of unified science projects or even the best evaluation for this particular project. Different projects will be faced with different criteria and different constraints; thus their evaluations must also differ. The example is, however, intended to serve as an illustration of one possible evaluation of a project of this type, and to clarify the nature of the output from each component process.

For example, the evaluator of the hypothetical project might include the following as some of the potential purposes for the evaluation:

(1) To provide formative information to the project staff concerning the development of a set of interdisciplinary science laboratories with respect to the feasibility, the practicality, and the instructional approach (inquiry versus post hoc verification) of these laboratories;

(2) To provide summative information to any interdisciplinary science program that might consider using these laboratories (a) regarding their cumulative effects on student approaches to interdisciplinary problems independently, and on student interest in science, and (b) regarding the specific setting and methods used in the pilot program with respect to laboratory space and apparatus, number of students, number and qualifications of teachers, length of time allotted, roles of the teacher and student, and pre- and post-laboratory preparations;

(3) To provide summative information to the funding agency regarding the processes used in the general development and evaluation of the laboratories, the development of each specific laboratory, generalizations about the success of each laboratory, and the cost of the process;

(4) To provide information on long-term effects to schools or school systems interested in using the laboratories regarding the retention of the learnings described in number (2) above;

(5) To provide the science education profession with summary information concerning the results of the total evaluation program.

Additional insight can be gained by placing these same purposes in outline form and including priorities indicating the relative importance to the project of each of the potential purposes. The format for the outlining is illustrated below:

Purpose	Priority	Sub-Priority
I. Type of information	(Ranked 1 to 5, where 1 is high priority)	
A. Decision-maker		
1. Focus		
a. Specific question		

The Roman numerals designate the type of information required. Capital letters designate the decision-maker for whom the information is to be obtained. The focus of each purpose is listed with Arabic numerals. Finally, the specific questions to be asked for each purpose are designated by lower case letters. The priorities are ranked one through five, with five being the lowest priority.

Purpose	Priority	Sub-Priority
I. Formative Information		
A. Project staff		
1. Development of interdisciplinary laboratories	1	
a. Feasibility		(3)
b. Practicability		(1)
c. Teaching technique		(2)
II. Summative Information		
A. Other unified science programs		
1. Cumulative effects of laboratories	2	
a. Student approach to investigate interdisciplinary problems		(2)
b. Student ability to investigate interdisciplinary problems		(3)
c. Student interest in science		(1)
2. Specific setting and methods	3	
a. Laboratory space and apparatus		(4)
b. Number and nature of students		(3)
c. Number and qualifications of teachers		(3)
d. Time allotted		(3)
e. Roles of students and teachers		(2)
f. Pre- and post-laboratory preparations		(2)

Purpose	Priority	Sub-Priority
B. Funding Agency		
1. Process used in laboratory development	1	(1)
a. General approach		
b. Each laboratory		
c. Generalizations with respect to success		
d. Cost		
2. Process used in evaluation	1	(1)
a. General approach		
b. Each laboratory		
c. Generalizations with respect to success		
d. Cost		
III. Long-Term Effects		
A. Other interested schools and school systems		
1. Retention of learning in all areas listed under II-A-1	3	
IV. Total Summary		
A. Science education profession		
1. Results of total evaluation program	4	

The stating of rationale for the assignment of priorities accomplishes two things. First, it diminishes the chance of priorities being assigned arbitrarily or on the basis of personal bias. Second, these rationale will be of help as the evaluator eliminates some of the potential purposes due to the constraints considered in the process of optimization.

Purpose	Rationale	Priority	Sub-Priority
I.	Very important to program development	1	
a.	Information available on similar programs		(3)
b.	Crucial to success of program		(1)
c.	Important for ultimate success of program		(2)

Purpose	Rationale	Priority	Sub-Priority
II. A 1.	Important for encouraging use of materials in other schools and determines extent that goals are met	2	
a.	A major goal of project		(2)
b.	A major goal of project		(3)
c.	Primary goal of project		(1)
II. A 2.	Important for interested schools to determine suitability and for replication	3	
a.	Not considered ideal and probably not duplicable		(4)
b.	Will give some indication of suitability for interested schools		(3)
c.	Will give some indication of suitability for interested schools		(3)
d.	May assist interested schools in planning		(3)
e.	Important as an illustrative model		(2)
f.	Important as an illustrative model		(2)
II. B 1.	Required by funding agency	1	
a. through d.	As required		(1)
II. B 2.	Required by funding agency	1	
a. through d.	As required		(1)
III.	Provides additional information on achievement of goals	3	
IV.	Will give professional recognition and promote unified science approach	4	

There are many potential purposes that would be appropriate for this particular project. The example identifies only a few of the purposes that the project evaluator might consider.

I.b. Identification of Constraints

Input

The input for the process of identifying constraints is grouped into constraining factors and mitigating factors. The project's financial situation, the availability, number, and qualifications of evaluation personnel (limitations on evaluation personnel), the present state of the project in terms of time, progress, and location, the ethics of evaluation, the political environment, and the availability of needed resources are some of the constraining factors that the evaluator must take into account. Mitigating factors, those which lessen the effects of the constraints, include sources of additional funds, volunteer assistance, opportunities for the evaluation personnel to improve their evaluative skills, evaluative information that may already be available, and the enhancing effects of good public relations (Figure 4).

Consideration of financial constraints basically consists of asking the question, "How much evaluation can the project afford?" If the evaluation is carefully planned before funding is sought, then much of this problem can be avoided by requesting an amount adequate for the needs of the evaluation desired. Even so, unexpected cutbacks or partial funding can introduce additional financial constraints. If, on the other hand, support for the evaluation comes from local sources, then the total amount available to the project may be more limited and may be allocated before the evaluation plans are fully formalized. Some of the budget items that must be considered include evaluation materials (tests and score sheets), computer time for scoring and analysis, reporting costs (determined by the form of the report), and salaries for the evaluator, assistants, secretarial help, and consultants. If the project

Ib. IDENTIFICATION OF CONSTRAINTS

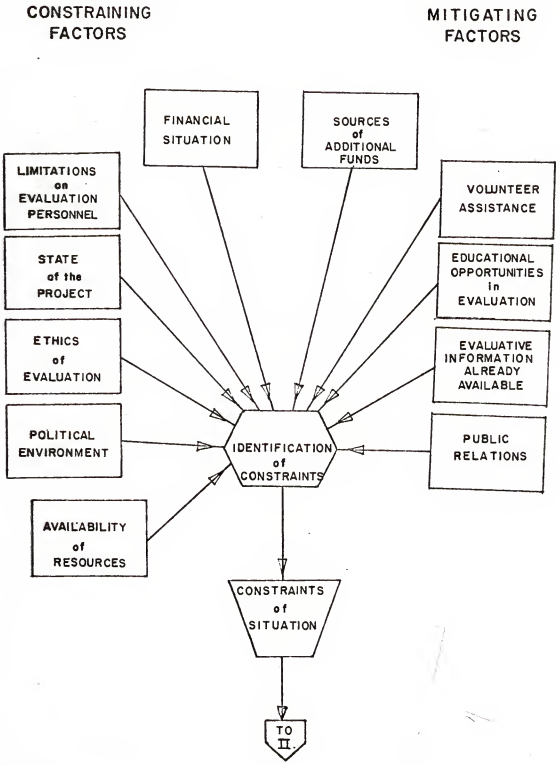


Figure 4.--Flow Chart for Identification of Constraints

involves several schools, then the cost of communication with these schools and transportation for visits to the schools must also be included. Other costs such as conferences for feedback purposes, instrument construction, and office space may also be important budget considerations.

A second constraint concerns the personnel designated as part of the evaluation staff. What is the staff allocation for evaluation? Is there an evaluation director, an evaluation team, a group of assistants, or consultant help? Are these people employed full-time, part-time, or are they expected to conduct the evaluation in their spare time? What other responsibilities do the members of the evaluation team have? Are they expected to bear responsibilities in other areas of the project (curriculum development, administration, public relations, management)? Finally, what qualifications do the members of the evaluation team have concerning both the process of evaluation and the subject area, unified science?

In the small-scale project, frequently only one or two people, employed part-time, are assigned to the task of evaluation. Often these evaluation personnel have had only minimal formal training in evaluation. Just as frequently, they are also expected to be full participants in the curriculum development process. In this situation, the evaluation must be limited due to the inability of one or two persons to handle an extensive evaluation. Multiple responsibilities can also produce conflict among roles. A person without the necessary qualifications will need more time and fewer additional responsibilities than would an experienced evaluator in order to carefully plan the evaluation.

Although an adequate number of evaluation personnel having sufficient time and excellent qualifications would be ideal, such a situation seldom exists. Due to personnel limitations many evaluators are faced with the choice of inadequate evaluation or no evaluation at all. In that situation, the evaluator must be aware of his limitations and modify the extent of the evaluation so that the evaluation attempted will be meaningful.

A related limitation is the amount of space available to the project evaluation. Evaluation requires careful filing, rapid access to files and to evaluation schedules, and work space. Without this, the evaluation can become much more time consuming and could be plagued by the losses of evaluative data.

Another critical constraint is the state of the project when the evaluation is begun. At what stage is the project, early planning, late planning, or implementation? The later in the project the evaluation begins, the more restricted the evaluation. For example, if evaluation begins late in the trial use of project materials, formative evaluation could be used for only part of the curriculum development. Any summative evaluation to determine change in student achievement or behavior would be severely limited due to lack of base-line data, that is, information about the subject's abilities, attitudes, and knowledge before exposure to the project materials. However, there is still much worthwhile evaluation that can occur if the constraints of the situation are understood.

Other questions pertinent to the state of the project are: Has adequate time been planned for analysis and reporting? Does enough time remain to develop new instruments that are needed or must existing ones

be used? How much of the school year remains? Does the physical location of the project limit the evaluation in any way?

The availability of needed resources such as computer services, consultant help, graduate assistants, schools willing to cooperate in the testing of project materials, and adequate library references on evaluative instruments and techniques can all contribute to the success of an evaluation.

The political environment is a constraint that is too often not considered. The primary question is whether the environment is supportive, friendly, indifferent, or hostile. A project that is considered out of the question at one time may be very popular several years later. Of great importance is the determination of the extent to which systematic evaluation is politically feasible. There can be two types of experimentation, replication or demonstration of the known, and exploration of the unknown. Funding agencies, school administrators, and other decision-makers are often intolerent of failure or insufficient success. When dealing with explorations of the unknown, there is a real chance of failure. The evaluator must know how much failure will be tolerated. If the environment is generally hostile, rather than jeopardize the entire project on the basis of the evaluation, it may be wise to evaluate it in parts.

In some cases, there may be constraints in the form of restrictive laws or policies, or even policies that would specify the evaluation methods to be used.

The final constraint considered here is ethics. This also is an important constraint that is seldom considered. The following list of questions suggests some of the ethical constraints that could arise. To

what extent is the evaluation an invasion of privacy? Is the data collected under false pretenses or without the subject's awareness? Is testing used for grading as well as evaluation? Who, other than the evaluators, will have access to the evaluative information? How much testing can you impose on students and teachers without producing negative effects or infringing on their rights?

In identifying constraints it is important to realize limitations and, on the basis of these limitations, reduce the extent of the evaluation to a size that can be realistically handled. However, there may also be ways in which the effects of the constraints can be lessened by mitigating factors. Location of sources of additional funds could certainly ease the financial constraints. There may also be people willing to donate time to assist evaluation for college credit, for experience, or simply because they see it as a worthwhile project. Sometimes the donation of office space can also be arranged. If the project is located at or near a college or university, the evaluators may have an opportunity to increase their competencies in the use of the evaluation process. Much evaluative information may have been already collected by school guidance personnel and may be available for evaluative studies. In some cases this could be a big savings in time and money. Good public relations and favorable press coverage may effect the political climate and reduce pressure on the project to demonstrate total success. These are only a few mitigating factors, but they are factors which could have a significant importance to the success of the evaluation. The creative evaluator and project director can probably find many more ways to circumvent the effects of constraints.

Process

The process of the identification of constraints consists of four steps.

1. The determination of the constraints that apply to the particular evaluation in question through a consideration of the list of possible constraints given above and through the inclusion of additional constraints as necessary.

2. The identification of any factors that could mitigate the effects of the constraints. These factors would be taken from the above list and from the imagination of the evaluator.

3. The final identification of the constraints that cannot be controlled. As the mitigating factors lessen or remove the effects of constraining factors, the list of effective constraints is shortened to a list of uncontrolled constraints.

4. Determination of the relative importance of the various constraints identified. This importance should be established in terms of the extent and seriousness of the effects of the constraints on the potential purposes for the evaluation.

As an example of this procedure, consider the hypothetical, example project developing and pilot testing interdisciplinary laboratory experiences and materials at a large, suburban public high school. The evaluator is a member of the science faculty who has been released part-time to work with the project. Among the constraints this evaluation faces are the part-time status and additional responsibilities of the evaluator, the nature of the school, a lack of clerical assistance, and the fact that the project was well underway before the evaluator joined the project.

By talking with professors at a nearby university, the evaluator is able to obtain the part-time assistance of a graduate student who would like to have the experience of working with the project and will receive credit for this experience. Through post hoc questionnaires and interviews and through the examination of the school records of the students, the evaluator is able to obtain information that reflects the nature of the project and the participants before he became associated with the project, thus reducing the effects of that constraint.

The remaining uncontrolled constraints which the evaluator is unable to alter are his own additional responsibilities and part-time status, and the nature of the school. He is also only partially able to remove the constraints of his late start.

Finally, the evaluator determines the possible relative importance of these constraints. The time constraints on this evaluator are probably very important in limiting the extent of the evaluation and should be carefully considered so that the quality of the evaluation is not sacrificed. The nature of the school limits the generalizability of the results and limits the evaluation purposes to those concerning this student population or similar ones. The late start could be crucial because of the loss of base line data for studies of change in student learning or attitudes. Any systematic formative evaluation of materials developed and tested would be hampered by the need for additional trials.

Output

The output from the process of identification of constraints is simply a listing of the uncontrolled constraints and their relative importance. For the hypothetical example project, such a list could take the form shown below:

Uncontrolled Constraints	Importance
1. Time limitations on evaluator	1. Limits amount of evaluation that can be successfully attempted.
2. Nature of project school	2. Limits generalizability and purposes of the project.
3. Late start of evaluation	3. Limits studies of change due to loss of base line data. Limits formative evaluation to untried materials or special retrials.

II Optimization

Input

The primary input into the optimization process are the potential purposes for the evaluation and the constraints of the situation. These input, along with the relative costs of the potential purposes, permit the evaluator to make a final decision about the purposes of the evaluation. An additional input concerning the project plans permits the evaluator to examine the relationships among the purposes and the relationships between the purposes and the project. (Figure 5.)

The criteria and constraint input have already been fully described as the outputs of the proceeding processes, I.a. and I.b. "Relative cost" refers to the cost in terms of money, time, effort, and effects on the other purposes. The cost in each case will vary with the purpose and with the way in which the evaluation is attempted. Since there is continued testing, analyzing, and recommending in formative evaluations, they are generally very expensive in terms of time and effort. Yet the monetary expenditure may be relatively small (with the exception of the evaluator's salary). Much summative evaluation requires extensive time and effort only near the beginning and end of a trial use of the program's materials. Such evaluation may be fairly

II. OPTIMIZATION

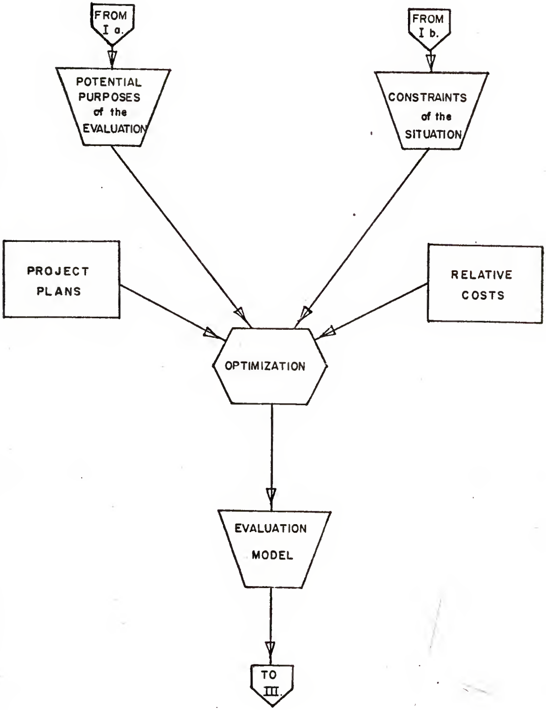


Figure 5.--Flow Chart for Optimization

expensive monetarily, however, due to the need for quantities of computer time for processing and analyzing of data. Studies of classroom transactions require sizable travel budgets, especially if several widely dispersed schools were involved in the project. Such studies involve numerous observations of each classroom. Consequently, they are expensive both in terms of time and effort. A long term evaluation requires that an evaluator be retained or rehired after the curriculum development has been completed. Since salaries are the major expense item in most evaluations, long-term evaluation would have to be considered very expensive monetarily. Cost, relative to the effects of one evaluation purpose on other purposes for evaluation, refers primarily to instances of conflicting purposes. There are other effects to be considered, as well. Extensive testing may produce a fatigue effect in student participants. The administration of one test may prepare a student for a test concerning another purpose. Significance of these costs will depend on the constraints of the situation. A project with little time, but sufficient funds, may not be concerned with monetary costs, but may find it necessary to make choices on the basis of time limitations.

The fact that an evaluation is costly does not mean that it is undesirable. Worthwhile evaluation is often expensive in several ways. When there are numerous constraints involved, choices are necessary. It is best, then, that these choices are made after a careful and realistic consideration of the possibilities and limitations.

The input on the operational plans of a project includes both the major stages of the project and a general time schedule for their implementation. There is no one way to adequately describe the operation of every developmental project. The identification of the major stages, however, may be facilitated by the consideration of the following

list of nine stages representative of those that are commonly included in curriculum development projects:

(1) Identification of needs
(2) Organization of project
(3) Establishment of project goals
(4) Development of project materials
(5) Trial use of materials (pilot testing)
(6) Revision of materials as necessary
(7) Second trial use of materials in additional schools (field testing)
(8) Second revision of materials
(9) Dissemination of materials and evaluation reports

The time schedule of the project should generally indicate the time of year and the year in which each of these stages begins and ends. The schedule for the Developmental Economic Education Program (DEEP) as described by Hulda Grobman in Developmental Curriculum Projects: Decision Points and Processes (Grobman, 1970, pp. 21-22) is an excellent example of such a schedule.

Process

The process of optimization consists of two phases. The first is the comparison of the criteria and constraints along with the cost of each purpose in order to determine which purposes will be finally accepted as the purposes of the evaluation. The comparison includes the determination of how each constraint will affect each specific potential purpose, identification of any purposes that might be affected by the constraints, weighing of the cost of each purpose against its priority, consideration of how the modification or elimination of various purposes would effect the constraints on other purposes, and finally, the summarization of the purposes so that the total evaluation will fall within the sum total of the constraints.

For the hypothetical example project, there are three major constraints. The time limitations on the evaluation personnel will mean

that not all of the evaluation identified can be undertaken. Some of the purposes outlined on page 48 will have to be modified or deleted. The nature of the project school will not be a limitation on these purposes since each of them can be accomplished successfully at this school. The generalizability of the results, however, will be limited as mentioned earlier. The late beginning of the evaluation process will impose some definite limitations. The summative purposes (II-A-1) are limited because of the lack of base line data about the student's abilities and interests before the beginning of the program. This does not mean that summative evaluation cannot be attempted. Much meaningful information can still be obtained either by using a design that does not require pre-testing or by administering pre-tests on the assumption that noticeable gain will still occur between this time and the completion of the program, even though several laboratories have already been performed. In either case, there are additional sources of error that must be taken into account during the analysis of the results. Formative evaluation is also affected. The laboratories that have already been performed cannot be properly evaluated except by their retrial with another group of students, an expensive and awkward procedure. Some information could be obtained about these laboratories by using post hoc questionnaires of interviews. There are, however, increased sources of error in such procedures. By looking over the outline of potential purposes, one can see quickly that all the purposes will be affected to some extent by the constraints.

After clarifying the effects of the constraints on the potential purposes, the cost of each purpose must be considered. A table, such

as the one prepared for the hypothetical example evaluation (Table 2), may be useful in comparing costs and making a final determination of the purposes of the evaluation. The two left-hand columns of the table refer to the priorities and the outline of potential purposes developed in the Identification of Criteria (page 48). The next four columns are for the cost of each purpose and are headed "Time," "Effort," "Money," "Effect." The final column is for an estimation of the relative total cost of each purpose. Additional space is provided on the right for the consideration of sub-purposes as needed. Spaces are also left beneath each cost estimate for notes. For the purpose of this illustration and since quantitative estimates of cost are difficult to determine, a relative scale is used ranking the cost as "Very High," "High," "Moderate," "Low," "Minimal," and "None."

Once a table has been constructed, the evaluator must consider the priorities, and within the limits of his constraints, make choices among the potential purposes. If all of the more costly purposes are also of low priority, then the decision is fairly simple. It is more likely, however, that the evaluator will be faced with questions such as whether a high priority, expensive purpose is of sufficient importance to warrant the exclusion of several purposes of lower priority and lower cost. As Hulda Grobman states, "Evaluation always reflects compromises. There are so many questions that could be asked, that not all can be answered satisfactorily." (Grobman, 1970, p. 179). The evaluator must continually recheck his criteria and constraints so that if the constraints change, he can make the necessary changes in the purposes of the evaluation.

Table 2.--Optimization of Evaluation Purposes

Purpose	Priority	Costs: Time	Effort	Money	Effect	Total	Notes and Sub-Purposes
I	1	High	High	Moderate	High	High	Expensive but Important a - Mod. - 3 b - High - 1 c - High - 2 DELETE a
					Everyday through entire project		
II.A.1.	2	Moderate	High	High	Moderate	High to Moderate	Expensive but Important a - Mod. - 2 b - V.H. - 3 c - Mod. - 1 DELETE b
				Time, material, and computer time			
II.A.2.	3	Low	Moderate	Minimal	Low	Low	Very Inexpensive ACCEPT
III.B.	1	Moderate	Moderate	Minimal	Low	Low	REQUIRED
			Record the processes in action				
III.	3	Moderate	High	High	Minimal	Moderate	Very Desirable But Not Crucial To Project Fairly Expensive DELETE!
				Must reemploy evaluator			
IV.	4	Low	Low	Minimal	Minimal	Low to Minimal	Low Priority But Very Inexpensive ACCEPT
		Summarize and report only	No additional information gathering	Only paper and typing			

For the example of the evaluation process that has been considered thus far, the decision is made to delete purpose III, "Long-Term Evaluation." This type of evaluation is highly desirable and important. But, unfortunately, the constraints placed on this project necessitate its deletion in order that more crucial questions can be considered. Purpose I, formative evaluation and purpose II-A-1, summative evaluation, must be modified due to the cost. The most economic and least disruptive modification can probably be accomplished by the deletion of I-A-1a, the feasibility study, and II-A-1b, the measure of student ability to solve interdisciplinary problems. The remaining purposes have been accepted as the "purposes of the evaluation." Other projects or evaluators, with other constraints, would probably make other decisions. But, at this point, these choices seem best for this project. The final purposes of the evaluation are as follows:

I. Formative information
    A. Project staff
        1. Development of interdisciplinary laboratories
            a. Practicability
            b. Teaching techniques

II. Summative information
    A. Other unified science programs
        1. Curriculum effects of laboratories
            a. Student approach to interdisciplinary problems
            b. Student interest in science
        2. Specific setting and methods
            a. Laboratory space and apparatus
            b. Number and nature of students
            c. Number and qualifications of teachers
            d. Time alloted
            e. Roles of students and teachers
            f. Pre- and post- laboratory preparations
    B. Funding agency
        1. Process used in laboratory development
            a. General approach
            b. Each laboratory
            c. Generalizations with regard to success
            d. Cost

2. Process used in evaluation
    a. General approach
    b. Each laboratory
    c. Generalizations with respect to success
    d. Cost

III. Total summary
  A. Science education profession
    1. Results of total evaluation program

The second phase of the optimization process is the determination of the relationships among the purposes of the evaluation and between those purposes and the project itself. One approach to this task is the construction of flow diagrams for a project's entire development-evaluation process. Figure 6 is a stylized representation of such a diagram.

The diagram consists of four primary elements essential to any description of the total development-evaluation process: (1) the various stages of the developmental curriculum project (Stage I, Stage II, Stage III, etc.), represented by the operations boxes (□); (2) the evaluation purposes, represented by the small, numbered circles (①); (3) the types of decisions or decision-makers involved, represented by the large circles identified by letter (Ⓐ); and (4) the effects that the decisions have on the project and on the environment, represented by dashed lines (--▷). In addition, the solid lines (——▷) represent the flow of the curriculum development process. The braces (}) represent the study and description of a process. The diamond-shaped figure (◇) represents the comparison of data from two or more sources. Finally, the dashed rectangle surrounding the entire process (⬚) represents the boundary of the project. Outside this boundary is the environment which consists in part of the society, the profession, and the community.

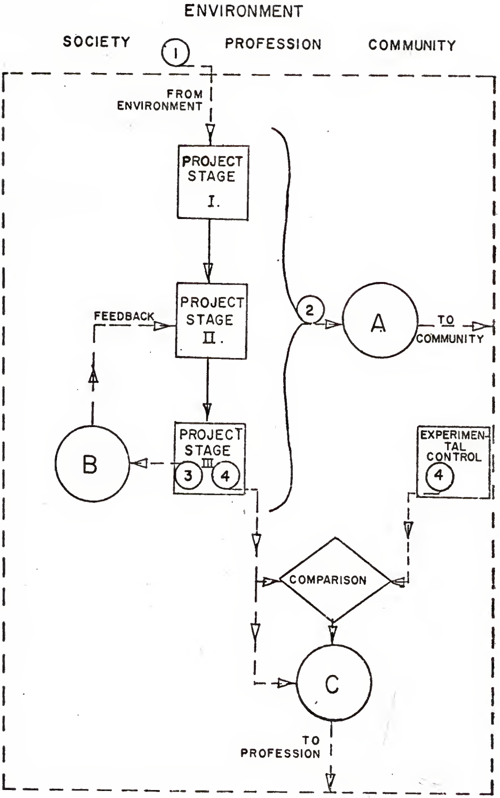


Figure 6.--Stylized Flow Diagram of the Development-Evaluation Process

The stylized representation begins with ①, an investigation of the environment. Information from this investigation is supplied directly to Stage I of the developmental curriculum project. One use for this kind of information would be to determine the goals of the project. Stage I leads into Stage II which, in turn, leads into Stage III. However, throughout these stages the process of curriculum development is being studied and described, ②. The information gained is reported to the decision-maker, (A), who with that information, makes decisions that affect the community. In Stage III, two purposes for evaluation are considered. Evaluation ③ provides information to the decision-maker (B). His decision, in turn, modified the program at Stage II. Another evaluation, ④, is conducted with the project materials at Stage III and with those of the control group. The results of these data collections are compared and then are sent to the decision-maker, (C). Information can also travel directly to (C) from the evaluation of the project materials ④. Decisions made by (C) will then have an effect on the science education profession.

In additon to using the general flow diagram presented above as a framework for the construction of a model for the development-evaluation process of a specific project, the evaluator may find it necessary, especially for complex processes, to consider various sub-processes separately. Formative evaluation, summative evaluation, and individual pupil monitoring are complex sub-processes that could require such consideration. The evaluation system of the Conceptually Oriented Program in Elementary Science, (COPES) (Shamos and Barnard, 1970, p. 3), is an excellent example of the representation of the formative purposes of an evaluation using a flow diagram. The COPES evaluation system

illustrates how questions of assimilability, feasibility, and learnability of the project materials are answered, who answers them, and how the answers effect the development of the materials in question.

Output

The output from the process of optimization should specify the final purposes of the project evaluation, the relationships among the purposes of the evaluation, and relationships between the purposes of the evaluation and the plans of the project.

The final purposes of the evaluation of the example project have been outlines in Process. A flow diagram of the development-evaluation process for this project is shown in Figure 7. According to this diagram, the project begins in the winter of school-year one with the identification of needs, the establishment of project goals, and the start of project operation. These lead directly to the development of materials beginning in the spring of school-year one and continuing through school-year two. As the materials are developed they are introduced into the school setting for trial use with both A (inquiry) and B (post hoc verification) methods of instruction. Information obtained by the formative evaluation (feasibility, purpose I-A-1a) of both A and B trials is reported to the project staff, which then modifies the materials on the basis of this evaluation. Other formative evaluation (teaching techniques, purpose I-A-1b) produces information on the two methods used to teach the project materials. This information is compared and the project staff determines the relative value of the two methods in teaching the various materials developed. During the same period of time, a summative evaluation of the school setting is undertaken (purpose II-A-2). The results of this evaluation is reported to

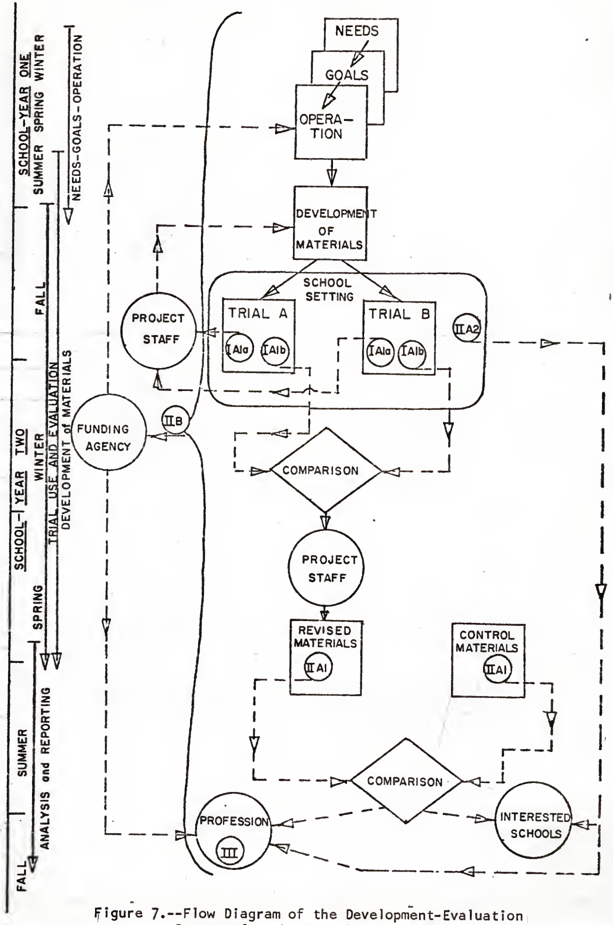


Figure 7.--Flow Diagram of the Development-Evaluation Process for the Example Project

other interested schools to assist them in deciding whether to adopt the project materials and to the science profession. The summative evaluation (purpose II-A-1) of the revised project materials is compared with the summative evaluation (purpose II-A-1) of the control materials during the summer of school-year two. These results are also reported to interested schools and to the science education profession. Descriptive evaluations of the curriculum development process (purpose II-B) continue throughout the existance of the project. The funding agency which requires this evaluation may make decisions based on the results which effect the operation of the project, the operation of other projects, or the science education profession. Finally, the results of all of the evaluations are collected into a summary (purpose III) and reported to the science education profession. Decisions made by members of the profession may effect the acceptance of the project materials by science teachers, the funding of related projects, or the future of unified science.

III Design Development

Input

The design development process, illustrated in Figure 8, seeks to specify the particular methods, instruments, and scheduling to be used in meeting the purposes of the evaluation. Consequently, in addition to the purposes of the evaluation and the evaluation model, a substantial amount of input is required. The evaluator must know from what sources he will obtain the evaluative information and also the nature of the evaluation environment. He must become aware of what alternatives are available to him in terms of data-gathering techniques and instruments and in terms of evaluative design and scheduling. In

III. DESIGN DEVELOPMENT

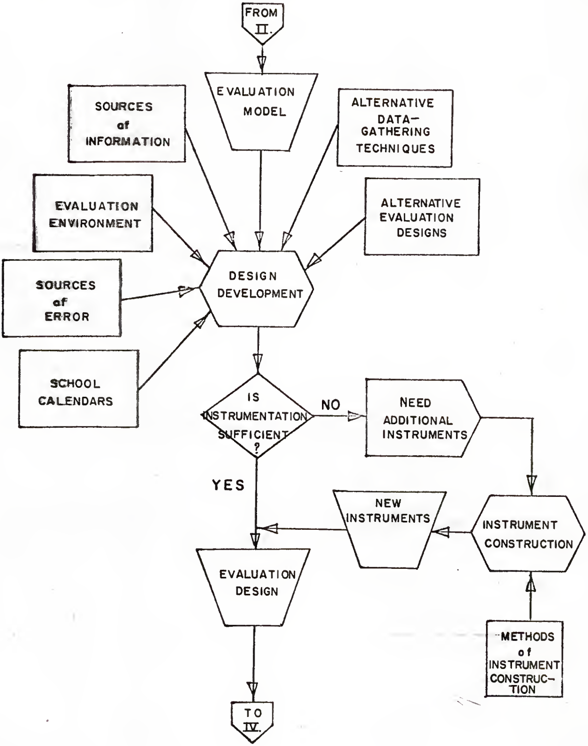


Figure 8.--Flow Chart for Design Development

addition, he should have some idea of the sources of error inherent in each alternative and information about the calenders of participating schools.

Each major purpose for evaluation implies the collection of information from a number of different sources. For the Assessment of the Social Setting, information could be obtained from the literature, members of the community, political leaders, administrative personnel, and project personnel. Context Assessment information sources would be science education literature, professional educators and scientists, and project personnel. For Antecedent Assessment, the evaluator might look to students, teachers, school records, the school administration, and the physical environment for information. In the Assessment of Developmental Procedures, the project records and project personnel would be suitable information sources. Studies of Transactions would require obtaining information from students, teachers, the physical environment, and project personnel. Information sources for Formative Evaluation include students, teachers, parents, and in some cases, the school administration and the physical environment. Students and teachers would be the primary information sources for Individual Pupil Monitoring. To identify Dysfunctional Outcomes, the students, teachers, and parents would probably be the best sources. Depending on the specific purposes of a Summative Evaluation, information might be obtained from students, teachers, school records, parents, administrative personnel, or project personnel. In Long-Term Evaluation, students, teachers, parents, school records, administrative personnel, and professional science educators would be called on for evaluative information.

Input on the nature of the evaluative environment concerns the number and locations of the schools involved, the number of students involved, the number of classes, the number of students in each class, the number of teachers and their teaching loads, the facilities available in each classroom, the length of the class periods, the time of day that the participating classes meet, the class schedule of each student, and certainly much more. This input will assist the evaluator in identifying any glaring inequalities among groups and in selecting an appropriate evaluative design.

A wealth of data-gathering techniques and instruments is available to the evaluator. There are so many choices that confusion often results. It is beyond the scope of this study to give a detailed listing or explanation of these techniques and instruments. However, a listing of the major categories of data-gathering techniques and a few examples of types of instruments used for such techniques might be of use to the evaluator in making decisions concerning the design of his evaluation.

Testing refers to a situation in which the subjects (information sources) respond to questions or problems and in which their answers can be judged for the degree of correctness. Tests can be written, oral, or performance and may involve words, pictures, situations, programmed devices, and so on. Tests are generally grouped into several different types of instruments. Some of these types are: standardized tests, including aptitude tests, achievement tests, intelligence tests, developmental tests, and readiness tests; criterion referenced tests; teacher (or staff) constructed tests; behavior tests; and situational tests.

The use of questionnaires is another common technique for data-gathering. Questionnaires are concerned with what a subject's answers to specific questions are, rather than the correctness of his answers. They may require unstructured comments, short answers, indication of degree of agreement, or simply agreement or disagreement. Examples of questionnaire instruments are: free-response questionnaires, opinion polls, temperament and adjustment inventories, interest inventories, rating scales, self-rating scales, and biographical data blanks.

A third technique of gathering data is the use of interviews. Interviews are direct discussions held for generally predetermined purposes. Interviews can be held with individuals or groups and may be structured, unstructured, or semistructured. Examples of types of interviews are diagnostic interviews, treatment interviews, and research interviews.

Observational techniques require a "...directed or intentional awareness or scrutiny of particular facts" (English and English, 1958, p. 353). The observation may be "on the scene" or may use some mode of recording such as audio or video tape recording. The observer may or may not be a participant in the observed activity. If the observer desires descriptive information about the setting, he may use a checklist type of instrument. In the event he desires only general, unstructured information, he may use an informal observation with informal recording of notes. Should he desire information on particular behaviors, there are a number of systematic observation instruments available that he could use.

Projective techniques differ from interviews and observations in that the situation is predetermined, but generally response is not

required. Projective techniques may make use of verbal, written, or performance responses. Some of the types of projective techniques include those using pictures, ink blots, situations, words or incomplete sentences to elicit responses.

The case study is a technique in which all available information about the individual's or a group's behavior and environment is collected and drawn together to present a total picture of a limited span of that individual or group's life.

Review techniques refer to the process of analyzing written material in order to determine characteristics of the materials. If the review concerns the materials developed by the project, then the reviewers may be students, teachers, subject area experts, educators, or others. They may be attempting to determine correctness, usefulness, and readability. Reviews of professional literature may also be conducted, especially by the project staff, in order to determine trends, needs, and methods.

The final data-gathering techniques considered here are non-reactive techniques. Such techniques are attempts to collect information without in any way disturbing the subject of the information. This information may come from volunteered comments, out-of-class behavior, attendance, and from many other sources if the evaluator is open and ready to receive it.

Evaluative design considers the questions of to whom data gathering procedures will be directed and when they will be used. Generally, such designs fall into six categories. Pre-experimental designs are those which attempt little or no control over the experimental variables. True experimental designs control these variables to a great

extent through the use of randomization techniques. When full control is lacking, expecially in natural settings, quasi-experimental designs are of most use. Though internal validity in such designs may not be as high as in the true experimental design, the external validity for these designs may be higher. Correlation designs are those which are used to establish correlation between variables, but not causality. Ex post facto designs are basically those in which data-collection occurs only after the experimental treatment and the subjects are then matched in order to obtain comparable groups. Finally, nonexperimental designs are those in which there is no treatment and the object is the careful description of a situation or process. The detailed description of the design in each of these categories is beyond the scope of this study. However, additional information can be found in Experimental and Quasi:- Experimental Designs for Research (Campbell and Stanley, 1963) and references to new designs can be found in the articles, "Choosing a Future: Strategies for Designing and Evaluating New Programs" (Light and Smith, 1970) and "A Design for Curriculum Evaluation" (Welch and Walburg, 1968).

A source of error is an uncontrolled variable which in some way effects the results of an evaluation. Error sources that are not accounted for generally lead to misinterpretation of the results. Consequently, some understanding of error sources is desirable.

Certain errors are inherent in each of the various techniques and instruments used to obtain evaluative information. In testing, the major errors probably have to do with the validity of the instrument. Validity concerns whether or not the instrument will measure the factor that we wish it to measure. Cultural biases and reading levels

introduce additional sources of error in testing. In the case of questionnaires, poorly asked questions (reliability) and sampling biases are major sources of error. By reliability is meant the dependability and accuracy of an instrument. The interviewer's sensitivity and ability to establish rapport, the physical setting, and the context and form of the questions are potential sources of error in interviews. A major source of error in the use of observational techniques is the reliability of the observers. In the use of projective techniques there is always a possibility that the person administrating the instrument will influence the subject in some way. The skills of the observer are possible error sources in case studies. But perhaps of more importance is the external validity of such studies (i.e., the extent to which the results are generalizable). In analyzing reviews, the evaluator must be concerned with the personal biases of the reviewer. Nonreactive techniques are least subject to error since the investigator does not impinge upon the life of the subject. Primary error sources in this case result from inaccurate recording and interpretation.

The various evaluative designs also are subject to inherent sources of error. The reference by Campbell and Stanley, mentioned earlier, gives a complete discussion of such error sources.

An additional input of use to the evaluator is the school calender and the calendars of any other subjects of the evaluation. Such information will help the evaluator to set up a schedule for data-gathering with a minimum of conflicts or disturbances.

Process

The process of development primarily consists of selecting among alternatives in the light of the evaluation environment, the school

calendar, the potential sources of error, and of course, the input from the proceeding process, the purposes and the model of the evaluation. In this process the evaluator must consider four questions. First, "What sources of information are available in this evaluation environment and which of those sources will provide the best information for each of the purposes of this evaluation?" For example, in the evaluation of the hypothetical interdisciplinary laboratories, the evaluator determines that for his formative purposes the most profitable sources of information will be the students and teachers using the project materials.

Second, when he selects the evaluative design, the evaluator is asking, "What arrangement of treatments (use of the experimental variable) and data collected can best provide meaningful information for each purpose under these conditions?" Here the evaluator must consider the evaluation environment (especially whether the groups are random collections) and the sources of error inherent in the evaluative designs as well as the alternative arrangements available. In the hypothetical example, the following decisions are made for the purpose of formative evaluation. Information collection before the use of the materials would be of little significance to this project. Observation of student behavior during the use of the materials would be very important. A significant amount of information could be obtained from students and teachers through questionnaires, interviews, and brief testing immediately after treatment. If this were carried out in both experimental groups (one using an inquiry approach and the other using *post hoc* verification), then the results could be compared to give some indication of the teaching methods best suited for use with these

materials. However, as a source of error, a major danger is that the two groups are not comparable, since they were not randomly assigned. The evaluator must find some way of dealing with this or consider the consequences of this source of error in his analysis of the results.

Third, the evaluator must also ask, "What data-gathering techniques and instrumentation will be necessary in order to obtain the information desired for each purpose?" Along with this question the evaluator must consider whether the needed techniques and instruments are available, or whether he will have to develop them. The process of instrument development is often a difficult and time consuming task. Since the chance of error is probably greater in project-made tests, special care must be taken in their creation. However, because there are so many possible purposes for evaluation and a wide variety of evaluative situations, there is frequently a lack of appropriate instrumentation and techniques. Consequently, the evaluator must decide whether he will construct new instruments and techniques or revise the purposes for the evaluation.

In the hypothetical example evaluation, the formative purposes are so specific that there are no appropriate instruments available for the observations, interviews, questionnaires, or testing. However, the evaluator determines that the purpose is sufficiently important to warrant the construction of any instruments necessary. Important to this decision is the consideration that assistance in the construction of these instruments might be obtained from the university nearby.

Finally, the evaluator must consider the question, "What schedule for evaluation activities should be followed to facilitate the complete and orderly collection of the information desired?" A schedule

must be appropriate for the evaluation design and must be considered in relation to the calendars of the schools involved.

In the hypothetical example, only one school and one calendar is involved. Consequently, scheduling of data-collection is simplified. For formative evaluation, data-collection occurs throughout the use of the materials (in this case, the entire year). The late beginning of the evaluation eliminates the possibility of gathering data on the first laboratories. From then on, evaluation will occur during and immediately following the use of each laboratory activity, beginning on the earliest possible date, October 21, and ending on May 8. By ending these evaluation activities early in May, allowance is made for the completion of the series requiring a longer time than anticipated and for avoiding the confusion generally associated with the end of the school year.

In addition to the activities described above, the evaluator should at this point give general consideration to the plans for analysis and reporting. These two processes will be more fully considered later in this chapter.

<u>Output</u>

The output from the process of <u>Design Development</u> should specify for each purpose: (1) the sources of information, (2) the evaluative design to be used, (3) the data-collection techniques and instruments to be used, and (4) the scheduling for the collection of information for each purpose. In addition, this output should also include a description of the evaluation environment.

The output for the hypothetical evaluation example used throughout this chapter is detailed below.

Evaluation environment.--A. There are two experimental classes of fifteen and eighteen students having one teacher using the inquiry method with the Interdisciplinary Science Laboratory Series.

B. There are two experimental classes of fifteen and twenty students having a second teacher using the post hoc verification method with the Interdisciplinary Science Laboratory Series.

C. There are two control classes of seventeen and twenty students having a third teacher using a chemistry laboratory series.

All students are in the eleventh grade at the same school. Course selection was optional with no randomization.

Evaluation design.--For purpose I-A-1a and I-A-1b (refer to purpose outline page 66):

(1) (sources of information) Data will be obtained from students and teachers of all experimental classes.

(2) (evaluative design) Observations will occur during treatment.

(3) (data-collection techniques) Teacher informal observations will occur during treatment.

Student questionnaires will be given after treatment.

Brief, project-developed, subjective tests will be administered after treatment.

Occasional student-evaluator interviews will be held after treatment.

Information from experimental group A will be compared with experimental Group B.

(4) (scheduling) This data-gathering will be conducted for each laboratory beginning with Laboratory #7 at a rate of about one per week from October 21 through May 8.

For purpose II-A-1a (refer to purpose outline page 66), a decision is made to pre-test even though delayed.

(1) (sources of information) Data will be obtained from all experimental and control students.

(2) (evaluative design) A nonequivalent control group design will be used as defined by Cambell and Stanley.

(3) (data-collection techniques) A project-developed, written-answer projection instrument will be used with the mixed administration of two forms.

(4) (scheduling) The pre-test will be administered to all experimental and control students on October 17 (one-half experimental and one-half control using form A, one-half experimental and one-half control using form B). The post-test will be given on May 24 with each student using the form he did not use in the pre-test.

For purpose II-A-1b (refer to purpose outline page 66):

(1) (sources of information) All students.

(2) (evaluative design) Same design as for II-A-1a.

(3) (data-collection techniques) A project-developed, semantic differential questionnaire with one form will be utilized.

(4) (scheduling) This will be administered on October 24, (preassessment) and on May 11, (post-assessment).

For purpose II-A-2 (refer to purpose outline page 66):

Purposes "a" and "d" will be obtained by one observation of the physical setting on November 7 using a checklist.

Purposes "b" and "c" will be obtained by one review of the school records on November 8.

Purpose "e" will be obtained by observation of students and teachers on five occasions: November 15, December 12, January 28, March 2, and April 17.

Purpose "f" will be obtained by review of project materials--a continuing process.

For purpose II-B (refer to purpose outline page 66):

(1) (sources of information) Data will be obtained from the project director.

(2) (evaluative design) This is a continuous process.

(3) (data-collection techniques) The general method for obtaining this information is by the observation and recording of the ongoing process.

(4) (scheduling) This is a continuous process.

For purpose III (refer to purpose outline page 67):

This is a summary report of all information obtained in the evaluation.

IV Implementation

Input

Input into the implementation process can be divided into three categories: input concerning the creation and collection of data; input concerning the management activities that must occur for successful data-collection; and input concerning the organization of the data collected. The first category includes the output from Design Development (evaluation design) and the techniques of building and maintaining rapport with the project, the participants, and the decision-makers (rapport building techniques). The second category includes logistical activities, scheduling techniques, appropriations

and cost of the evaluation, and the budgeting techniques. The third category consists of organization techniques for storage and rapid retrieval of data (Figure 9).

The evaluation design specifies the steps to be taken in the creation and collection of data. It outlines the particular instruments and techniques to be used, lists the persons who will participate in the evaluation, and designates the time and location of each data collection activity. The maintenance of rapport is also an important activity for an evaluator. The lack of rapport with participants could lead to major biases in the results or could alter the willingness of the participants to cooperate to the extent necessary for adequate data-collection. A loss of rapport with the project staff or with the decision-makers involved could lead to their being unwilling to accept the results of the evaluation.

There are many techniques that might be of use in establishing and maintaining rapport with the participants. When a pilot study of a field test is beginning, the evaluator might contact the administrators of each of the participating schools to discuss the project and evaluation with them and if possible find ways of easing the load on the teachers involved. If sufficient funds are available the project could offer the participating teachers a compensation for their time spent working on the project. Often a teacher will harbor fears that the evaluative results concerning him and his class will be used as an indicator of his abilities as a teacher. A sensitivity to the feelings of participating teachers and, whenever possible,a strong assurance of confidentiality may help assuage these fears. As the evaluation progresses, individual and group interviews with students, teachers, and

IV. IMPLEMENTATION

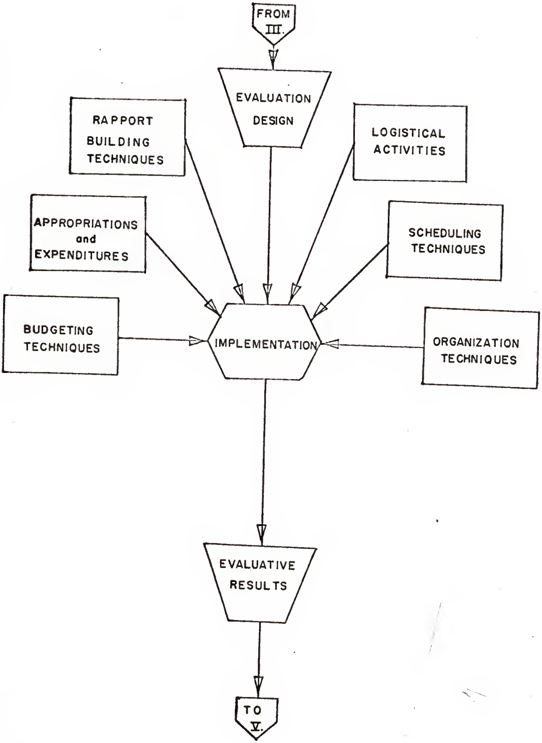


Figure 9.--Flow Chart for Implementation

administrators concerning their veiws of the evaluation and concerning the evaluation's effect on them will help air complaints and increase the willingness of participants to undergo the evaluation. However, this may also increase the error due to an increase in the Hawthorne Effect. Whenever possible, the involvement of participating teachers in the actual curriculum development would also help maintain rapport. Forcing a teacher's cooperation by administrative decree or not listening to the advice about project materials given by teachers will lead to many problems. Even seemingly insignificant actions may be important, such as supplying stamped, self-addressed envelopes for the return of evaluative information.

Maintenance of rapport with the project staff or with decision-makers concerned is equally important. Regular, but informal, conversations with project personnel will keep them in touch with the progress of the evaluation effort. Careful wording of evaluation reports so that they are not threatening to the people involved will increase the willingness of the project personnel to accept the results of the evaluation.

To assure the smooth conduct of the data-collection process many management activities are necessary to support the creation and collection of data. If the evaluation is using tests or questionnaires, then they must either be obtained from commercial sources, or if they are project-produced instruments, then they must be reproduced. It may also be necessary to obtain or produce score sheets. The correct number of instruments must then be distributed to each of the participating schools by the time required by each different school calendar. Instructions are of great importance. The incorrect use of

instruments or variations in the way they are used or the conditions of their use can introduce major sources of error into the evaluation. Instructions should be clearly written, stating the precise procedures to be used and the specific instructions to be given to the students. Whenever possible, additional verbal explanations with opportunities to ask questions about the instructions would be advisable. The evaluator must also make some provision to assure the prompt return of the evaluative information. The use of observational or interview techniques as part of the data-collection may require special training for the person who will use these techniques. If the evaluation requires the use of certain equipment such as audio or video tape recorders, arrangements should be made well in advance of their planned use. If the participating schools are scattered through more than one town, then the evaluator must prepare for rapid and frequent communication. This will require correct addresses, telephone numbers, and prearranged times at which the participating teachers can easily be reached by telephone. Visitations to such schools must be scheduled in advance and transportation arranged. If feedback-conferences are planned then the arrangements for them must include meeting rooms, accommodations, and meals. Finally, the evaluator must continually monitor these tasks to make certain that they are properly carried out.

Many management activities will occur long before the actual data-collection begins. For each of these activities, however, timing is important. Late distribution of evaluative instruments, for example, may negate the possibility of their proper use. Consequently, there is a need for precise scheduling of all data-creation and collection activities and logistical operations. This scheduling should include:

the kinds of activities; the amount of time, material, and personnel required; the approximate starting and completion dates for each activity; an indication of how the activities relate to one another and to the rest of the project; and a priority assignment that indicates the importance placed on meeting the scheduled deadline for each activity. "In one sense this provides a job description for all the people connected with evaluation; it explains the breakdown of the budget... and it serves to explain who is doing what, when, and why." (Grobman, 1968, p. 89). One currently popular scheduling system that is useful for this purpose is the Program Evaluation and Review Technique (PERT) (Dresdner, Speich, and Uslan, 1963).

An additional input concerning management tasks is a budget, the amount of money allocated to each of the various evaluation activities and an accurate recording of expenditures connected with these activities. Closely associated is input concerning budgeting techniques. A last minute shortage of funds to purchase the appropriate evaluation materials or to pay the cost of scheduling visitations can be a serious deterrent to the success of the project evaluation. It is beyond the scope of this study to provide extensive information concerning budgeting techniques. However, one source of information who is usually readily available and who generally has up-to-date information about local budgeting procedures is the school business manager. Consultation with this person could ease the operation of the entire project.

The organization of raw data requires a method for its storage and retrieval and a technique for its coding. Methods of storage include the use of file cabinets or similar storage facilities, the use of

audio or video tapes, the use of microfilm, and the use of digital computers. Coding of data for storage and rapid retrieval can be based on several considerations. Data can be coded according to the evaluative technique, the evaluative instrument and form, the source of the information, (school, teacher, class, student), the treatment, the date of data-collection, the use intended for the information obtained, and the probable analysis technique that will be used.

Process

The aim of the implementation process is to provide meaningful evaluative results. Meeting this aim requires the use of three subprocesses: (1) the creation and collection of data and the maintenance of rapport; (2) the performance of management activities; and (3) the organization of raw data.

The creation and collection of data refers to the administration of evaluative instruments or the use of evaluative techniques by the evaluator, and the production of information by the subject of the investigation relative to his unique history and his current environment. Basically, this is accomplished by carefully following the procedures outlined in the evaluation design. However, in conjunction with this the evaluator must identify ways of maintaining rapport with with the teacher and student participants, the project staff, and the decision-makers. The maintenance of rapport might be aided by the consideration of the rapport techniques listed in Implementation, Input, and the identification of others that would specifically apply to the project in question.

The performance of management activities requires the identification of the logistical operations necessary for the evaluation, the

scheduling of these activities, and the continuous appraisal of the financial situation. For example, in the hypothetical evaluation only project-produced instruments are used. No instruments need to be purchased. However, plans must be made for the reproduction of instruments. In this case two thousand three hundred copies of the basic questionnaire form, one hundred fifty copies of each form of the projection instrument, and one hundred fifty copies of the semantic differential instrument are mimeographed in September. Formative tests for each laboratory are reproduced on ditto two days prior to each laboratory session, eighty copies each time. Written instructions for the use of each instrument are given to each teacher the day prior to its use. The same day brief meetings are held to review the instructions and answer questions. Since all of the participating classes are in a single school, distribution is not a major problem. To reduce the danger of loss, arrangements are made to distribute the instruments on the day they are to be used. The data from the formative instruments are kept together by the teacher until immediately after the use of all of the evaluative instruments for a single laboratory has been completed. The data from the projective instrument and the semantic differential instrument are collected immediately following their use. The observations for formative evaluation are to be informal and conducted by the participating teachers, consequently no training will be attempted. The evaluators will conduct all interviews and have already had training in interview techniques. Communication between the evaluators and participants in this project does not present a serious problem, since informal meetings can occur frequently; however, they are not guaranteed. Consequently, to increase the opportunity for communication concerning

the project evaluation, arrangements are made for brief monthly meetings of the evaluators and all participating teachers.

Next, each of these activities must be scheduled along with other evaluation activities. Figure 10 shows a small segment of such a schedule for the sample evaluation.

Figure 10 is a PERT network for the events that make up the formative evaluation of laboratory number seven and the beginning of the evaluation of laboratory number eight. The numbered circles in the diagram represent "events," noteworthy points in the evaluation which denote the beginning or completion of a task. An event requires neither time nor resources. The solid arrows connecting events represent "activities," the actual performances of tasks. These do consume time and require resources such as manpower, material, space, and facilities. The dotted arrows are "dummy activities." They indicate that the next events occur immediately without additional expenditures of time and resources. Below the activity lines are listed the person(s) responsible for the accomplishment of the task indicated. The following is a keyed list of the events scheduled on this diagram.

Events

1. Begin Preparation of Formative Instruments
2. Begin Implementation of Formative Evaluation
3. Begin Pre-Laboratory Class Preparation Seven
4. Begin Instructions for Teachers
5. Complete Aquisition of Tape Recorders and Tapes
6. Begin Laboratory Seven
7. Complete Laboratory Seven
8. Begin Teacher Observation Seven
9. Complete Teacher Observation Seven
10. Complete Post-Laboratory Activities Seven
11. Complete Distribution of Test Seven
12. Complete Administration of Test Seven
13. Complete Writing of Test Twelve
14. Complete Distribution of Questionnaire Seven
15. Complete Use of Questionnaire Seven
16. Complete Interview

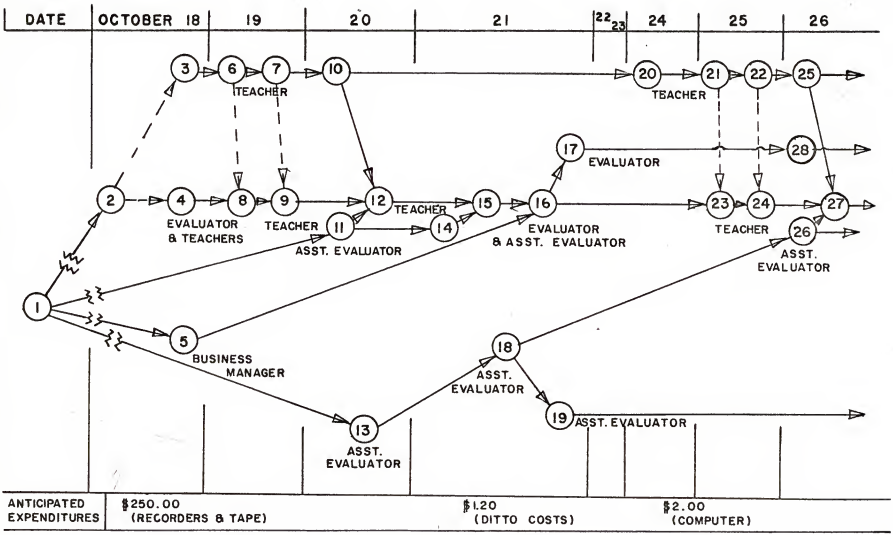


Figure 10.--PERT Network for Formative Evaluation of Laboratory Number Seven

17. Complete Collection of Data on Laboratory Seven
18. Complete Reproduction of Test Eight
19. Begin Writing of Test Thirteen
20. Begin Pre-Laboratory Class Preparation Eight
21. Begin Laboratory Eight
22. Complete Laboratory Eight
23. Begin Teacher Observation Eight
24. Complete Teacher Observation Eight
25. Complete Post-Laboratory Activities Eight
26. Complete Distribution of Test Eight
27. Complete Administration of Test Eight
28. Complete Analysis of Formative Data for Seven

If estimates are made of the "optimistic" and "pessimistic" times required for the completion of an activity, and if additional PERT techniques are applied, the evaluator will be able to determine the extent to which the evaluation is conforming to the schedule. He can identify potential areas in which insufficient time or resources may endanger chances of the project's finishing on schedule. These techniques can also aid in the establishment of priorities and the allocation of resources.

The appraisal of the financial situation for evaluation activities refers to the process of comparing the initial budget and anticipated cost of each activity with the actual expenditures on a continuing basis. This might best be accomplished by posting a record of these budget considerations on the chart of scheduled activities so that an estimate of remaining costs will be readily visible.

At the bottom of the diagram is a row of figures identifying expenditures anticipated for various tasks. Additional rows could indicate other budget information such as the actual expenditures and the total amount of funds remaining at any particular time. Consultation with the school business manager would probably be helpful in accomplishing this task.

The organization of data "...includes providing a format for classifying information and designating means of coding, organizing, storing, and retrieving the information," (Stufflebeam, 1969, p. 72). This is accomplished by identifying the storage methods that are most suitable for the data and developing a coding system that will facilitate the rapid retrieval of data in accordance with the needs of the project evaluation. For the hypothetical project evaluation, all written data are placed in filing cabinets. The only other data are taped interviews which are shelved until analyzed, and then reused or retained as desired.

A data code may consist of any number of indicator digits and the coding will vary according to the evaluation purpose for which the data were collected. For the hypothetical evaluation, the code established for formative and summative data uses five indicator digits. The outline number of the evaluative purpose (refer to page 66) is the first indicator and for clarity is placed in parentheses. The second indicator is a single letter representing the technique used to obtain this data (e.g., T = test; O = observation; I = interview; P = projective instrument; S = semantic differential instrument; and so on). The third indicator identifies the treatment and class from which the data were collected (e.g., X_{A1} = experimental class A, section 1; X_{B1} = experimental class B section 1; C_2 = control class, section 2). For formative data, the fourth indicator is the laboratory number for the data. For summative data, however, the laboratory number is not necessary. Instead, the fourth indicator tells whether the data are from pre-treatment or post-treatment administrations of the evaluative instruments, and which form of the instrument was used (e.g., 1_A = pre-evaluation with

with Form A; 2_B = post-evaluation with Form B). Finally, the fifth indicator identifies the student who produced this data. Thus, (l-la), T, X_{12}, 09, 087 would mean that the data identified by this code number were collected for the formative evaluation of practicality from the second section of the experimental classes using inquiry techniques, that it concerns laboratory number nine, and that it was produced by student number eighty-seven.

Output

The output of the implementation process is a collection of coded and stored raw data. In the hypothetical example evaluation, the formative data on each laboratory used in the experimental classes include teacher-recorded observations of the laboratory activities, answers from each student to a written questionnaire, and answers to a written test for each student. Data for the summative purposes include written replies from the pre-treatment and post-treatment administrations of the projective instrument and the semantic differential instrument for each student in both experimental and control, a checklist of physical characteristics for each class, a listing of all participants along with information concerning each one, written descriptions of the roles of students and teachers in each class, and written reviews of the finished materials. The results of summative evaluation for the use by funding agency will be in the form of a collection of daily written observations concerning the laboratory development and the evaluation procedures.

V. Analysis

Input

Analysis is the process of translating evaluative results into meaningful conclusions. The input for this process includes: (1) the results of the process of implementation; (2) the information needs of the decision-makers concerned; (3) the nature of the variables; (4) the possible levels of data; (5) the methods available for data analysis; (6) assistance available to aid in analysis procedures; and (7) when appropriate, additional comparative data.(Figure 11).

The information needs of the decision-makers which serve as input for the analysis process are basically the same needs that are considered during the identification of criteria. However, their reconsideration as part of the analysis has a slightly different emphasis. Originally the emphasis was to establish the purposes of the evaluation. Here, it is to determine how data collected for these purposes can best be analyzed to obtain the information desired by the decision-makers.

For analysis purposes, information needed can be grouped roughly into two categories, descriptive and inferential. Descriptive information describes the characteristics of the data obtained, such as the range, central tendency, and variability. Inferential information concerns "...inferences made from the actual data at hand to larger situations" (Fox, 1969, p. 236).

The planning for analysis must begin along with the selection of instruments for data collection and the establishment of a schedule. However, the flexibility required for a developmental curriculum project demands that these plans be open to change.

V. ANALYSIS

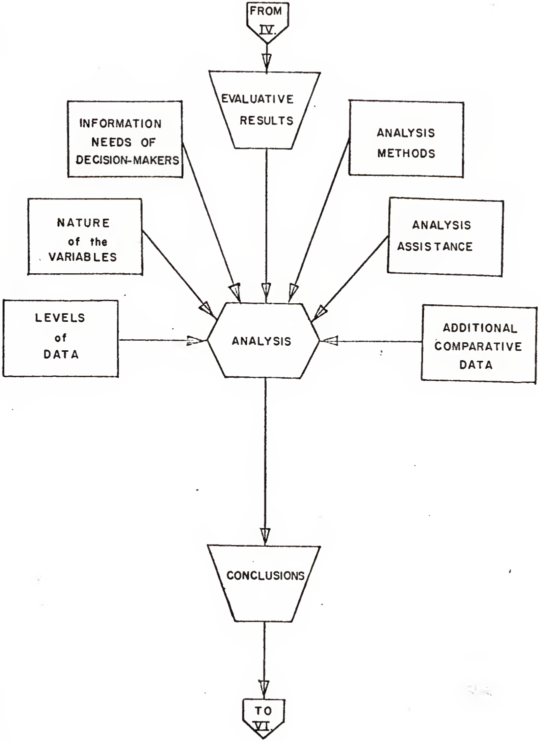


Figure 11.--Flow Chart for Analysis

A shift in the importance of the various decision questions facing a project, a change of the person who occupies the position of a decision-maker, or an alteration in a decision-maker's opinions can significantly affect information needs. The evaluator usually has little control over these variables. Consequently, he must continually be alert to such changes and ready to make adjustments for them.

The nature of the variables involved in the evaluation will also affect the analysis. Variables can be either discrete or continuous. Discrete variables are those which can be described only in whole units. Number of people, for example, would be a discrete variable. However, according to David Fox, discrete variables may be classed as: dichotomous, only two alternatives possible; limited category, three to six gradations; multiple category, seven to nineteen gradations; or infinite category, at least twenty gradations with no upper limit. (Fox, 1969, p.p. 139-140). For continuous variables, on the other hand, fractional units may exist. Time is an example of a continuous variable.

Another source of input is the nature of the raw data. Data can be either qualitative or quantitative and can exist at several different levels: nominal, ordinal, interval, and ratio. Nominal data refers to nonquantitative information concerning nonquantitative discrete variables. Ordinal data, indicating the relative position of an observation with every other observation, can be categorized as either verbal or numerical. Verbal ordinal data indicates position by using terms such as bad, poor, fair, satisfactory, good, very good, and excellent. An example of numerical ordinal data is ranking, that is, the assignment of a numerical order. However, there is no indication

of the amount of separation between ordinal classifications. Interval data provide this information with a numerical scale having equal gradations. Ratio data differ from interval data only in that the ratio data include an absolute zero, the point representing absence of the variable under consideration.

The fifth input concerns the methods of data analysis. Basically, there are five approaches to data analysis. The first approach, the complete restating of all evaluative data in a comprehensive form, implies only a processing activity with little or no true analysis. The second approach is the identification of the elements of evaluative data and their summation into more compact communication. This is primarily a processing function, but with some analysis. The analysis may indicate semantic content (the elements of the data), or feeling tone (the attitude communicated), or provide a basis for inference. Quantitative analysis may involve the use of basic sorting procedures, tallying and crosstallying.

The third approach, the tabulation of data, is an important part of both processing and analysis. By tabulating data relationships, trends and differences become more obvious, and crude predictions can be made.

A high level approach to data analysis, often ignored in educational research and evaluation, is the fourth approach, the use of graphing techniques. Frequently recommended as a method of displaying evaluative conclusions, graphing can also be a powerful analytical tool by giving a visual impression of the way in which two variables depend on one another. With this tool, relationships, trends, and differences are readily obvious, and predictions can be made with a fair degree of certainty.

By far, the most powerful approach to analysis of evaluative data, the fifth one, is the use of mathematical techniques. The educational researcher or evaluator must deal with large numbers of subjects that are mostly unpredictable when considered individually. Consequently, a statistical approach to mathematical analysis is most profitable for educational purposes.

Analysis of nominal data for descriptive purposes may make use of frequency distributions, the determination of the mode, chi-square, and phi coefficient. Depending on their level, other data can be analyzed using commulative frequency distribution, determination of range, the percentile system, and the moment system which includes variance, standard deviation, and standard error of the mean. The statistical procedures listed above can be used to provide information concerning the parameters of the data and the population. Additional descriptive techniques can be used to establish association (chi square, calculation of the contigency coefficient, the Q test, and calculation of the phi coefficient) or correlation (the Spearman rank-order correlation, Kendall's tau, the Pearson product moment correlation, the correlation ratio, tetrachoric correlation, biserial correlation, and point biserial correlation). When more than two variables are involved partial correlation or a correlational matrix must be used. To identify the underlying structure of interrelationships, factor analysis can be used. Finally, simple regression, multiple regression, and several other techniques are suitable for the prediction of the functioning of future groups.

Inferential statistics permits the evaluator to determine the probability that observed differences are attributable to chance and to

make consistent and objective predictions concerning larger situations from the analysis of his data. Essential to inferential statistics is the calculation of the significance probability, that is, the probability that observed differences could be due to chance variations. Techniques that facilitate these calculations are classed as either parametric or nonparametric. Parametric techniques generally are based on the assumptions that the sample characteristics exist in the population, that they are normally distributed there, and that the sample statistic provides an estimate of the parameter. Parametric techniques include the use of the normal distribution, the t distribution, and t-test, the F distribution and F-test, the analysis of variance, and the analysis of covariance. These techniques are limited to use with moment system descriptive statistics.

Nonparametric techniques make no assumptions about populations or their parameters. Consequently, they are appropriate for use with ordinal data and some nominal data and are applicable to small samples. However, nonparametric techniques are not as powerful as parametric techniques in rejecting null hypotheses when they should be rejected. Nonparametric techniques include application of probability theory, the binomial expansion, and the chi square distribution. (Fox, 1969).

Two additional types of analysis are important in determining the precision and accuracy of the measuring instrument and the techniques used in data collection and in data analysis. They are reliability, an indication of the extent to which repeated use of an instrument or technique will produce the same results, and validity, an indication of the extent to which a procedure accomplishes its purposes. Some of the techniques available for determining reliability are

alternate-form reliability, split-half or odd-even reliability, test-retest reliability, and Kuder-Richardson reliability. The effects of unreliability on data can be estimated by calculation of the standard error of measurement and by obtaining the correlation between variables. According to Fox, some of the kinds of validity are face validity, content validity, construct validity, concurrent and congruent validity, and predictive validity. (Fox, 1969).

Input concerning the amount of assistance available to aid in analysis procedures is important for determining the extent of analysis that can be practically undertaken by a particular development project evaluation. This assistance could include the availability of computer services, desk calculators, consultants on statistical analysis, statistics textbooks, and professional analysis services.

In some cases, additional input concerning comparative data will also be of help. Data on similar programs, data on the population under consideration, or data concerning the evaluation sample may be available to the project evaluator for a small expenditure of effort. Comparison of these data with the project data may substantially increase the value of an evaluation.

Process

The process of analysis consists of four parts: (1) the selection of analysis methods and techniques; (2) the anticipation of possible outcomes; (3) the processing of data; and (4) the statistical analysis of data. The selection of methods to analyze the data produced through the implementation process actually begins as the data collection instruments and techniques are identified. According to Fox:

> There are varied systems available for both descriptive and inferential statistics. The choice among them is determined by the nature of the variable being studied, how it is treated in the particular study, and the level of data which is then obtained. Thus there is relatively little freedom involved in the choice of a statistical technique. (Fox, 1969, p. 68).

As an example of the process of selecting statistical techniques for data analysis, in the case of the summative evaluation of the hypothetical example project, data obtained with the semantic differential instrument can be viewed as continuous, multi-category, interval data. The resulting information is intended to give interested schools an indication of the effects of the project on student interest in science. Because of the nature of the data, analysis can include the comparison of several different concepts on the basis of the means and variances for individual groups, of scores on specific scales (good-bad, hot-cold, etc.), total scores for a single concept, or sub-totals on major factors that combine several sub-scales. Graphical representations of "factor scores" can be produced using Osgood's concept of a "semantic space," an n-dimensional plot of the average scores for each concept on orthogonal dimensions (independent factors). (Sax, 1968, p.p. 270-271).

The use of the t-test and F-test can indicate significant differences between means and variances,respectively. The index of distance, D, can be calculated to indicate differences between profiles (Sax, 1968, p. 272). The use of factor analysis aids in the identification of major factors common to several scales. Since the three classes are not equivalent, analysis of covariance can be used to equate the groups and determine the significance of change from pre- to post-tests and of differences in gains among the three groups. However, these classes were not randomly selected. Consequently, nonparametric techniques paralleling those listed above should be used.

One further consideration is a choice that the evaluator does have, that is, the extent of the analysis. Researchers are expected to extract every bit of information possible from their data. However, in a developmental curriculum evaluation this may not be practical due to the constraints of time and money within which the evaluation must operate. In the example evaluation a computer and consultant help are nearby; consequently each of the techniques suggested will be used.

When possible, the description of potential outcomes in advance of data analysis permits more objectivity in the explanation of whatever outcomes do result. It also permits the collection of additional data that might be needed to explain the results.

Once planning is completed and the data has been collected, the processing and analysis of data consists of applying the calculations specified by the analysis techniques selected for use with the evaluation. Along with this, however, the evaluator must also be alert to unanticipated outcomes, such as relationships between sex or class meeting time and the treatment variable, and to information that is exceptionally useful, such as particularly insightful responses to questionnaires. In addition, the evaluator must be aware of, and when necessary, attempt to explain sources of error, such as, the use of statistical techniques based on assumptions that are not met by the evaluation, and the reliability, validity, and sensitivity of the instruments used in data collection and analysis.

Output

In the case of qualitative data, the output from the process of analysis consists of conclusions about the frequency and diversity of responses concerning what was communicated, the tone of the communication, and its underlying meaning. Statistical analysis produces

conclusions describing the characteristics of variables, the relationships among variables, the predictability of relationships among variables, the differences among groups with respect to a variable, the amount of change in a variable with respect to time, and the probability that the observed differences are due to chance (significance).

For example, in the hypothetical evaluation, the results of the analysis of the data collected with the semantic differential instrument are expressed as mean scores and variances for each of the concept areas. They also indicate the amount and significance of the gain in each of these areas. In this case, all groups show significant gains in the aspect of student interest in science identified as "Doing Laboratories." The gain, in this aspect of student interest in science for the experimental group using the inquiry approach, is significantly greater than the gains shown by the other two groups. But the differences in gains between the other two groups is not significant. The interpretation of the meaning of these conclusions is a part of the process of reporting.

VI. Reporting

Input

The primary input for the process of reporting is the collection of conclusions developed in the analysis process. Additional input must include: (1) the results of related studies; (2) the characteristics of the decision-making audience; (3) the needs of the decision-making audience; and (4) available methods of reporting. (Figure 12).

Results of studies similar to or related to the evaluation under consideration can be used to provide support for the evaluation findings, or to locate points not considered or in disagreement with the previous studies. Once these points have been identified,

VI. REPORTING

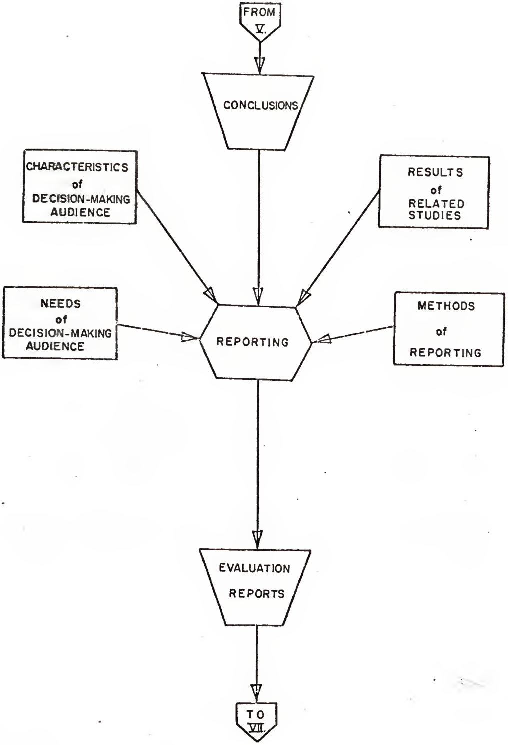


Figure 12.--Flow Chart for Reporting

comparison of differences in approach and conditions can lead to important insights into the evaluation. Identification of these studies is accomplished by a review of the related literature both at the beginning of the project and repeated periodically throughout its existence.

The characteristics of the decision-making audience refers to the level of sophistication of the decision-maker with respect to evaluation, science, and unified science, the decision-maker's openness to the idea of evaluative information aiding in the decision-making process, and the pressures and influences on him due to his decision-making position. The needs of the decision-making audience include the need for information as specified by the purposes of the evaluation, the need for the timely reporting of information, and the need for conciseness or extensiveness in the report of evaluative information.

Since the characteristics and needs of the decision-making audience will vary, so should the methods used for reporting. This includes the extent of the report, the level of the language used in the report, the structure of the report, and the medium used for its communication. A report can be all inclusive or limited to various portions of the evaluation. Its language can vary from informal language understood by all, to formal scientific language understood only by professional evaluators. The report's structure can range from that considered to be journalistic to the highly structured form of a "research report." The media available to provide a means of communicating the report include written formal reports, abstracts, articles for professional journals, newspaper articles, formal oral presentations, casual discussions, slide presentations, films, video or audio taped reports, and audio-visual aides such as transparencies and charts.

Process

The purpose of reporting is to translate statistical conclusions into information that will have meaning to the decision-maker for whom it is intended. This implies that not all decision-makers should receive the same report. In fact, an evaluation report, like any other communication, should be tailored to fit the needs, desires, and background of the intended audience.

The process of reporting includes three steps: (1) the interpretation of the conclusions; (2) the selection of reporting methods to be used with each decision-making audience; and (3) the preparation of evaluation reports.

In interpreting the conclusions of statistical analysis, the first step is an attempt to summarize them into more easily handled bits of information. This may involve explaining the meaning of the conclusions in language that can be easily understood by the decision-makers. Decision-makers vary in sophistication with respect to evaluation terminology and procedures. Consequently, the language of reporting must vary. The summarization must also emphasize points that are important to the purposes of the evaluation.

The next step is to find alternative explanations for the conclusions. Ideally, these explanations should be identified before data analysis begins so that they will be as objective as possible. However, explanations are frequently overlooked, or unanticipated results are found. In these cases the evaluator must attempt to identify all possible explanations for each result while taking care to minimize the effects of his own biases. The suggestions of a consultant not associated with the project would be useful.

Once alternative explanations are identified, the evaluator must select from them the explanation that seems most appropriate. In doing this, he must consider the factors that placed limitations on his evaluation, and consider the arrangements made to reduce the effects of these limitations. For example, if in the hypothetical evaluation, statistical analysis shows that group A's gains between pre- and post-testing are higher than the gains of group B, something produced this difference. There are a number of factors that might account for the differences, totally or in part. Chance factors are controlled by determining if the differences are statistically significant. If they are significant, then the evaluator must consider the possibility that factors other than the treatment variable were responsible. Such factors include: the biases of the evaluator; invalid, unreliable, or biased measurement devices; the research conditions, class meeting time, size, teacher, and so on; the conditions and methods of implementation; and the effects of the evaluation itself, the Hawthorne effect, testing effect, and fatigue.

In a careful research, many of the effects of identifiable limitations and uncontrolled variables will be specifically reduced in some way or controlled statistically, so that the treatment is the major factor. However, it is unlikely that the effects of all of these extraneous factors can be sufficiently reduced. In such cases, the factors must be considered in the interpretation, either as explanations for undesirable conclusions, or as cautions about the generalizability of desirable conclusions. Most evaluations take place in settings that do not permit the adequate control of the sample population or the control of many other variables. This is considered a serious limitation by the researcher. It also increases the difficulty of interpreting conclusions.

However, there are serious questions about the desirability of establishing highly controlled conditions for curriculum development evaluation. Perhaps, since curricula must be used in natural settings, intervening variables should be included in the evaluation (Guba, 1969, p. 34). In such cases, an alternative to the control of variables would be to fully describe the sample under consideration and any other identifiable variables. This would permit those who wish to apply the evaluative conclusions to their own situations to determine for themselves the similarity between their situations and that of the evaluation.

After appropriate explanations have been selected, then the evaluator is in a position to become more subjective in suggesting the implications of the evaluation for each decision. It is important, however, for the evaluator to remember that the decision-maker may not see the same implications. For example, the evaluator will probably accept a statistically significant difference as highly probable evidence of a casual relationship, and in some cases, as evidence that the treatment caused the difference. But, this is not necessarily an indication of the importance of that information to the decision-maker. Whether or not the difference is statistically significant, he is faced with the same question of "practical significance," "Is the change under consideration worth the gamble involved?"

As the evaluator interprets statistical evidence, he must also determine what methods of reporting will be most suitable for each decision-making audience. These methods of reporting are selected from the available methods in accordance with several considerations: the sophistication of the decision-maker with respect to his understanding of the terminology of evaluation, the evaluation process, and the meaning

of the statistical techniques used. These considerations will greatly effect the style of the report and the amount of original data included.

The evaluator must also consider the effectiveness of each reporting medium for each decision-maker. For example, newspaper articles would not be an effective means of reporting to a funding agency, but they could be very effective for reporting to the lay public. The expectations of the decision-maker for the evaluation results and for the form of the report should also be considered. For example, a report of negative results to the project staff would probably take a very different form from a report of negative results to the funding agency. In addition, this selection must also include a consideration of the skills of the evaluator in each medium and the amount of time and money available for the preparation of the report.

The next step in reporting consists of the preparation of the report itself. Depending on the media and the form selected, the evaluator will be faced with a variety of tasks. Assistance from consultants, professionals, or from the school's public relations and media personnel might be necessary in areas of communication unfamiliar to the evaluator.

Finally, the report is communicated to the decision-maker at a time that is suitable to his needs. An evaluation report after the decision has been made will be of little use to the decision-maker. Similarly, an evaluation reported too early is apt to be lost or forgotten when a decision is finally reached.

Output

The output of the reporting process is a report tailored for each of the decision-makers involved. A single, overall report is not sufficient. The report should include as much information as the

decision-maker needs and will be able to use, in the form and at a time most useful to him. It is often advisable, when possible, for the evaluator to accompany the report to the decision-maker and discuss with him the implications of the report. To suggest in the report what the decision-maker's final decision should be, is presumptuous on the part of the evaluator. This is not a part of the evaluator's role in the evaluation process.

For the hypothetical example evaluation, the report of formative information to the project staff consists of very informal, brief statements concerning evidence of areas of deficiency and strength. The reports are presented in writing after each laboratory trial; and oral explanations are given at weekly meetings. Reports of summative information designed for use by other interested schools are in the form of informal descriptive pamphlets with pictures. At the same time, however, the pamphlet summarizes completely the summative information. The funding agency requires a lengthy, formal report of the type usually described as a research report. The outline for such a paper includes: the purpose, a review of related literature, the hypotheses, a statement of procedure, data and results, and conclusions and implications. For reporting to the science education profession, several reports covering different aspects of the evaluation are needed. The reports for the hypothetical evaluation include articles for professional journals and papers to be presented at regional and national professional meetings. The presented papers include written copies and abstracts, oral presentations with questioning periods, and charts and graphs prepared on transparencies for overhead projection.

VII. Decision-Making

Input

Decision-making is the final component of the process of evaluation and the only one which is not in some way controlled by the evaluator. Input into the process of decision-making includes: (1) the identity of the decision-maker; (2) the nature of the decision; (3) the decision-rules; (4) additional information concerning the decision that is available to the decision-maker; and (5) internal and external pressures on the decision-maker that could affect the decision made. (Figure 13).

The identity of the decision-maker and the nature of the decision are closely related. The relationships between the two were discussed in Identification of Criteria. Since the decision-maker is the key figure in the decision-making process, his identity (i.e., his personality and his values) will probably be the most influencial factor in the decision. The nature of the decision to some extent dictates the choices available to the decision-maker. A developmental decision may determine whether to accept, reject, or modify the part of the curriculum in question. It may also determine what modifications, if any, are to be made. Operational decisions are more complex in that the number of alternatives is greater. Decisions might, for example, concern the scheduling of evaluative activities. An administrative decision involves the costs and commitments of the project to the school or school system in which it operates. The administrative decision-maker could have the project terminated if he felt that the costs were too great or that commitments were not adequately met. At the policy level, the decision might be to terminate a project if its goals are in conflict with the goals of the

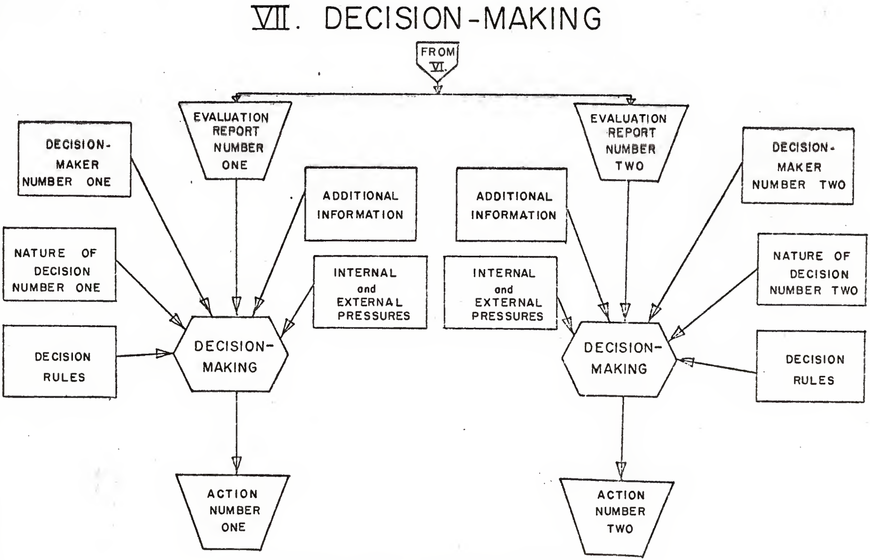


Figure 13.--Flow Chart for Decision-Making

community or society; or pressure might be brought on the project to force a change in its goals. Decisions concerning initial and continued support of the project with funds, space, time, and opinion are support decisions. Professional decisions would include whether to accept, accept partially, challenge, or reject the results of the project evaluation. They might also include consideration of how the preparation of teachers should be modified according to the evaluation results. A publisher's determining whether or not to publish the new material would be a dissemination decision. Finally, decisions of adoption are made by the representatives of interested schools concerning the extent to which the program is applicable to that school or how it might be modified so that it would be more applicable.

Decision rules are specifications of the criteria to be applied in determining which alternative decisions will be made. According to Garlie Forehand: "Information collected for decision-making will be useless unless someone has specified what outcome leads to what action." (Forehand, 1970, p. 29). If the decision-maker is a member of the project staff, the specification of decision rules is crucial. However, if he is not, it may be difficult to obtain such rules. The decision-maker may not know what decision-rules he will use. If he does, he may be unwilling to state them. If he states decision-rules, they may be unreasonable, too lenient, or reasonable in the opinion of the evaluator. By getting to know the decision-maker, the evaluator may be able to anticipate the decision criteria that will be used. He may also be able to assist the decision-maker in determining decision-rules or in modifying those rules. In any case, the evaluator does not have the prerogative of making the decision rules for the decision-maker.

In addition to the evaluation report and the explanations of the evaluator, the decision-maker may also have other sources of information concerning the decision that are unavailable to the evaluator. The decision-maker's past experience, his own evaluations concerning the decisions, and his knowledge of overriding considerations such as cost or legal barriers will, without doubt, effect his decision. The extent to which they effect the decision will depend upon the relative importance he places on each of them and on the evaluation report, and the extent to which he views each of them as valid sources of information about the decision. The project evaluation may play only a small role in the final decision, especially if the evaluator has not attempted to insure the decision-maker's understanding of the process of evaluation.

Finally, there are probably a number of internal and external pressures operating on the decision-maker that will influence the decision. Decisions are not made in isolation. The decision-maker's mental and physical condition, and his recent experiences are an inescapable part of the decision process, as are the needs and desires of those associated with him. The evaluator has little control over these factors, but awareness of them can help him in aiding the decision-maker in understanding the evaluation report.

Process

The process of decision-making includes three phases: (1) the presentation of the reports; (2) the assistance to the decision-maker in understanding the report; and (3) the making of the decision. The nature of the presentation is determined by the nature of the report, the identity of the decision-maker, and the nature of the decision. Since there are several types of decisions to be made and several different

decision-makers are involved, there must be a separate report and presentation for each one. Also of importance is the timeliness of the presentation, so that it will coincide with the decision-maker's readiness to make the necessary decision. It is also advisable that the evaluator be present to provide assistance in understanding the report and to answer any questions that may arise.

Generally, the evaluator is not a participant in the actual process of deciding. The decision-maker makes the necessary decisions within the framework of the information he has at hand, the internal and external pressures acting on him, the decision rules, if they have been stated, the nature of the decision, and his own biases. The process of decision-making is the subject of much study and has been the subject of many recent books and articles. A further discussion of the process is beyond the scope of this study.

Output

The output of the decision-making process consists of specific directions for action. If the person identified as the decision-maker is indeed the "effective" decision-maker, rather than a figurehead, then action will occur implementing the decision. For example, as the evaluation of the hypothetical project progresses, many formative decisions are made concerning modifications of the materials. Also, a decision is made to concentrate on the further development of materials based on the inquiry approach and to encourage the use of that approach with the materials.

The decisions by interested schools based on the summative evaluation of the project result in the adoption of the materials by several schools. Some schools reject the materials on the basis of cost or the

unsuitability of the materials for their particular situations. No schools elect to modify the materials.

Information supplied to the funding agency aids them in making decisions concerning the curriculum development and evaluation processes that they will recommend to future projects.

Finally, the decisions made by professional science educators based on the summary of evaluative information play an important role in increasing interest in unified science and in the inception of several additional projects concerning the development of unified science materials.

Summary

This chapter has considered the elements that make up the input, process, and output of each of the components of the process of developmental curriculum project evaluation. These components, Identification of Criteria, Identification of Constraints, Optimization, Design Development, Implementation, Analysis, Reporting, and Decision-Making, overlap both in time and function, and yet, are sufficiently separable to be useful constructs in understanding this complex process. While the discussion in general has been theoretical, examples concerning a single, hypothetical, developmental curriculum project evaluation have been offered throughout the chapter to give the discussion a practical viewpoint, as well. Although the hypothetical example evaluation is part of a unified science project, the picture of the evaluation process presented in this chapter has not been limited strictly to unified science applications. The following chapter, however, examines the implications of this model for projects concerned specifically with unified science.

CHAPTER V

USE OF THE EVALUATION MODEL IN UNIFIED SCIENCE

The small-scale, local unified science project is in several respects unique among curriculum development projects. First, the goals of unified science programs generally differ from those of "disciplinary" science courses. For this discussion, a "disciplinary" science course is defined as one of the usual biology, chemistry, or physics courses currently taught in secondary schools. Second, the nature of the curriculum frequently differs in content, approach, and length from disciplinary courses. Third, the impetus for each program has generally come from a few individuals actively teaching science, who feel that present approaches are inadequate. The motivation is internal rather than external (felt needs--rather than a new textbook or a federally funded national curriculum project). Unified science programs are primarily supported by local funds, universities, and small grants from federal and private sources. There is no nationwide unified science project.

Consequently, in using the evaluation model developed in this study and explained in Chapter IV, the evaluators of small-scale, local unified science programs must be aware of the unique qualities of these programs and how these qualities could affect the evaluation. The purpose of this chapter is to present guidelines for the use of the evaluation model in a unified science curriculum development project. In

addition, a bibliography is included to provide further references to additional information concerning the evaluation process.

Throughout the evaluation of a curriculum development project the process of each component (i.e., Identification of Constraints and Identification of Criteria) is basically the same for any project. However, the input varies significantly with the project evaluation.

In the Identification of Criteria, while the alternative "purposes for evaluation" available to the evaluator of a unified science project are identical to those available to any evaluator, the basis on which the purposes of a particular evaluation are chosen may differ. Input concerning the influences due to societal and personal biases are primarily dependent upon the environment in which the evaluation takes place. The project's impetuses and goals, and the decision questions which the project must face,are a function of the nature of the project itself.

An examination of the materials produced by twenty unified science programs and three articles concerning unified science resulted in a compilation of impetuses, goals, and characteristics that apply to the various unified science programs. Such compilations are sometimes dangerous because of the possibility of misinterpretation. These particular compilations do not represent the "ideal" unified science program, nor do they represent any single program. They are listed simply as illustrations of some of the ways in which unified science programs are and are not unique.

The impetuses that led to the development of various unified science programs are grouped into eight categories. Under each category heading is a brief summary of statements concerning the impetus for unified science development.

Some of the impetuses leading to the development of unified science programs are:

I. The interdisciplinary nature of nature
   A. Nature is multidimensional.
   B. Life and its problems make no distinction among disciplines.
   C. Scientific research is now frequently interdisciplinary.

II. The unity of science
   A. The three major disciplinary headings are obsolete and no longer clearly separable.
   B. There is excessive overlap in high school science courses.
   C. A fundamental goal of science is to simplify.
   D. The science disciplines are intellectual conveniences to facilitate specialized study.

III. Universal science literacy
   A. General education should be the primary function of high school science teaching.
   B. Science literacy is needed if students are to make intelligent decisions in our highly technological society.
   C. A common ground is needed where humanists and scientists can communicate.
   D. Most students now have little or no exposure to chemistry, physics, or the earth sciences.

IV. The interdependency of science and society
   A. Our science and society cannot exist independently. Each requires the support of the other.
   B. Students need to understand the interrelationships between science and the social, economic, and physical conditions of man's life.
   C. Science teaching should stimulate students to effect change to improve the quality of life for mankind.

V. The rapid accumulation of knowledge
   A. There is now too much science knowledge for any one person to teach or learn.
   B. Duplication of teaching must be limited to that essential for understanding.
   C. Content must be limited; and that to be retained must be selected on a rational basis.

VI. The nature of learning
   A. "Facts" are not the essence of science. They are neither lasting nor reliable.
   B. The concepts of one discipline often cannot be understood fully without background from another discipline.

C. There needs to be a logical content development in science.
D. The new and exciting topics of the other disciplines must often be excluded.
E. Science needs to be taught as a coherent whole rather than piecemeal, to facilitate conceptualizations of interrelationships.

VII. The importance of the individual learner
The student's interest and level of intellectual development should be of equal importance as the science in shaping the curriculum.

VIII. Attitude toward science
There are now many negative attitudes toward science activities, scientists, and the process of science in general.

From the impetuses summarized above, unified science programs have developed with a wide diversity of goals. Some of these goals are considered under seven topic headings below.

The goals of the twenty unified science programs surveyed include:

I. Knowledge of content
A. Understanding of selected facts, concepts, theories and generalizations, and increased skill in using them.
B. Acquisition of knowledge which can be used to explain, predict, understand, and control natural phenomena.
C. Preparation of science-oriented students for more rigorous science undertakings.

II. Knowledge of process
A. Recognition that the meaning of science depends upon its inquiry processes as much as on its conceptual schemes.
B. Development of an accurate image of science as an approach to experience and of the relationship between it and approaches characteristic of human action in general.
C. Understanding of the point of view and activities of a scientist.
D. Understanding of the tentative nature of scientific data and conclusions, and the importance of suspended judgment.
E. Understanding of general laboratory techniques and skills and methods of reporting.
F. Awareness of professional opportunities in science.

III. Ability in the appropriate use of processes of science
   A. Recognition of the application of scientific inquiry outside of science.
   B. Development of creative skills.
   C. Development of skills in problem solving through scientific inquiry.
   D. Ability to engage in the processes of science and to apply these processes in appropriate everyday situations.

IV. Understanding of interrelationships
   A. Recognition of the essential unity of nature.
   B. Understanding and appreciation for the interrelationships among the traditional science disciplines, and the unity of science.
   C. Awareness of similarities and differences between science and other endeavors such as technology, humanities, social sciences, philosophy, and religion.
   D. Insight into the formative forces (past, present, and future) acting reciprocally between science and other facets of society and culture.
   E. Appreciation for the role of science in a modern world.
   F. Better understanding of one's self, his relationship to others, and to his natural and social environment.

V. Scientific attitudes and behavior
   A. An accurate image of science as an approach to experience and of the relationship between it and approaches characteristic of human action in general.
   B. Understanding that science is one, but not the only, way of viewing natural phenomena, and that even among the sciences there are rival points of view.
   C. Appreciation for the limitations of science.
   D. Aquisition of the attitudes of scientists and the appropriate application of these attitudes in daily experiences.
   E. Ability to weigh advances and to accept those felt to be of value for the good of society.
   F. Aquisition of desirable behavior patterns including curiosity--open-mindedness--rationality--growth in science--patience--presistence--problem solving ability.

VI. Interest and enthusiasm
   A. Gain intellectual satisfaction.
   B. A lasting enjoyment and appreciation for science.
   C. Sympathy toward scientific enterprises.
   D. Increased participation in scientific activities.
   E. Interest and enthusiasm in science or technology as a hobby or vocation.

VII. Psychomotor skills
   A. Development of skills in the perception of objects, qualities, and relations.
   B. Adeptness in the use of various tools of measurement.
   C. Compentency in the use of laboratory techniques and skills.

Although the goals are not significantly different from the general goals of science education, the uniqueness of the goals of unified science is that they generally include <u>all</u> of the goals of general education for several or all disciplines of science. This breadth is unique to unified science.

The resulting unified science programs can be described by characteristics included in the list below. Each of the characteristics listed is a characteristic of at least one unified science program. These characteristics have been grouped according to focus, target population, role of the teacher, content, materials, and course structure and methods. Since this list is a compilation of the characteristics of many unified science programs, each developed to meet the particular needs of a single school or school system, some of the characteristics listed may be contradictory. The characteristics include:

I. Focus of curriculum:
   A. Unifying concepts or conceptual schemes
   B. Interdisciplinary themes
   C. Issues
   D. Traditional disciplinary concepts
   E. Processes of science
   F. Relationships of science and society
   G. Laboratory activities

II. Target population:
   A. Nonscience-oriented students
   B. Science-oriented students
   C. Low-ability students
   D. High-ability students
   E. All students

III. Role of the teacher:
   A. Diagnosing and prescribing
   B. Guiding

C. Moderating discussions
D. Lecturing (traditional role)
E. Team planning and teaching
F. Clarifying interdisciplinary ties

IV. Content:
A. Traditional disciplinary content--sequenced for understanding and efficiency
B. Traditional disciplinary content with many interdisciplinary modifications
C. Partially traditional content, partially content not from traditional disciplines
D. Complete departure from traditional content

V. Materials:
A. Traditional texts
B. Locally developed unified science texts
C. Articles, booklets, and miscellaneous materials
D. Multiple texts
E. Multiple audiovisual materials
F. Programmed materials
G. Laboratory materials

VI. Course structure and methods:
A. Inquiry approach laboratories
B. Open-ended laboratories
C. Multiple resource utilization
D. Behavioral objectives
E. Individualization
F. Self-pacing
G. Limited repetition of concepts
H. Reuse of previously learned concepts
I. Concept placement appropriate for ability and preparation
J. Logical content development
K. Careful sequencing of topic-issues-concepts
L. Inquiry-problem solving laboratories
M. Several-year sequence
N. Building from the general to the more specialized
O. All general-building to more sophistication
P. Interchangeable instructional units
Q. Flexible (used as supplement or as course)
R. Correlated laboratories
S. Interdisciplinary laboratories
T. Programmed and computer-assisted instruction
U. Mixed grade levels

The above list should not be considered a characterization of unified science courses. It is, instead, a list of characteristics which, in part, describes at least one of the twenty unified science programs surveyed.

Criteria for the evaluation of unified science curriculum development projects are identified by applying the process outlined in Chapter IV. The first step involves identification of decision questions important to the project. Developmental decisions necessary for a unified science project are determined by considering the goals and characteristics of the particular project. For example, a developmental decision might determine whether or not certain materials need modification, based on the extent to which they contribute to the goal of "understanding of interrelationships." Operational decisions are likely to be important to unified science projects due to the constraints of their small scale. Some of these constraints and possible approaches to circumventing their effects will be discussed later in this chapter. Administrative decisions may be especially important. The unified science project is often located in a single school or school system, consequently, the project will have immediate effects on the school and will probably be held responsible for those effects. Formal policy decisions affecting unified science projects will differ little from those affecting other projects. On the other hand, informal policy decisions, such as parental response, may differ due to the radical departure from tradition by many unified science projects and the lack of national prestige. Support decisions are important for all projects. For small-scale projects, however, many of the support decisions will be made on the local level. This presents both unique problems and unique opportunities for the evaluator. Professional decisions and dissemination decisions are likely to center on unified science's departure from tradition and on the worth and advisability of such a departure. Finally, the goals and characteristics of the particular project will be of primary importance in identifying adoption decision questions.

Unified science projects could be affected by any of the decision-makers considered in Chapter IV. However, in small-scale projects, local decision-makers are likely to play a more important role. This increases the opportunity for direct communication with the decision-maker, an opportunity that is not readily available to evaluators of many large-scale projects. In identifying local decision-makers, it is important to consider the possibility of two levels of decision-making, an official level and an informal level. Frequently, the informal decision-maker (the "man behind the scenes") is the more powerful. An understanding of the informal decision-making process is important to the success of small-scale, local projects. References on the informal decision-making process are listed later in this chapter.

The informational needs of the various decision-makers must either be determined by the evaluator and the decision-maker in cooperation, or be inferred by the evaluator on the basis of the personal characteristics of the decision-maker, the nature of the decision question, and the nature of the project being evaluated.

In theory, the potential purposes available for the evaluation of unified science projects are the same ones that are available to any curriculum development project. In terms of practical availability (i.e., the chance that a purpose will actually be selected) unified science projects have some advantages. Studies of the long-term effects of teaching are frequently advocated but seldom undertaken. Many unified science programs, however, span two, three, or four years. Consequently, long-term goals can be meaningfully examined with relative ease as an integral part of the project evaluation. This is also an excellent opportunity for sequential studies to determine the best arrangement of interdisciplinary learnings.

The uniquenesses of the societal values and personal values that influence the determination of evaluative purposes are primarily a function of the small scale of unified science projects. Rather than differences in the nature of the values, these uniquenesses exist mostly in the amount of influence that these values have on the evaluation of the project. Since small-scale projects are often intended to meet local needs, it is important that the evaluation be more responsive to community values and more attentive to the needs of local decision-makers, than are most large-scale projects. The evaluator of a small-scale project is likely to also be very much involved in the curriculum development phase of the project. This means that the evaluator is more likely to be familiar with the aims and methods of the project than a professional evaluator would be. However, this same interest in the project can be a major source of error, especially in identifying decision-makers and in setting priorities for the evaluative purposes. In this situation extra care must be taken to maintain objectivity.

Unified science goals are very broad; and many of the possible purposes for evaluation seem applicable. The constraints, however, of the small-scale project will probably limit the extent of the evaluation that can be undertaken without sacrificing the quality of the evaluation. The evaluator must carefully establish priorities among potential purposes so that the final purposes of the evaluation can be selected on a sound basis. The meaningfulness and usefulness of the evaluative information is more important than the extent of evaluation attempted.

In considering the constraints of the situation on the evaluative activities of small-scale, local unified science programs, one must first realize that these programs are faced with many serious constraints.

Understanding the nature of these constraints and their potential effects on an evaluation is an important step in planning the evaluation. Only if constraints are recognized, can an evaluator deal with them. Once they are recognized, however, he may, in many cases, be able to determine ways of diminishing or completely circumventing the effects of various constraints. If the evaluator can find no way to eliminate all constraints, then he must be able to adjust his evaluation to operate successfully within the remaining constraints.

Most of the constraints that must be considered in evaluating small-scale, local unified science projects are related to the small scale of the project. The financial situation of the project is likely to be more dependent upon the availability of local funds than would be the finances of large-scale projects. If a budget can be set up in advance, however, the locally funded project may have a higher degree of certainty of receiving all of the funds allocated to it than would the large-scale project. Even though limited, if the extent of funding is known in advance, plans can be made to operate within the necessary budget; or ways can be found to extend the budget. Often local businesses or individuals are willing to donate materials to worthwhile causes. Office supplies, the use of a calculator and typewriter, or the loan of an audio tape recorder might be obtained in this way. Local donations to any phase of the project, however, could release funds for evaluative activities.

Limitations placed on evaluative personnel are another major constraint for unified science projects. Due to the limited size of the project the evaluator may also have curriculum development responsibilities. This can introduce many biases into the evaluation. Some of these biases

can be overcome by using multiple data collection techniques. This approach will be explained later in the chapter. Frequently, in small-scale projects, the evaluator also has teaching responsibilities. If these responsibilities include the teaching of classes participating in the evaluation, then additional biases may be produced. Some of the effects of these constraints might be reduced by using data collection techniques that are not as susceptible to evaluator bias. The performance of data collection activities by someone other than the evaluator would be another way of limiting these effects. In addition, biases should be acknowledged in the interpretation of conclusions; and their possible effects and any attempts made to limit those effects should be explained.

It is unreasonable to expect the evaluator of every small-scale, local project to have had the experience and educational background in evaluation expected of the evaluator of national projects. First, there is probably not an adequate number of professional evaluators available, especially evaluators with experience in unified science. Second, in order to increase the number of evaluations attempted, it is important that more curriculum developers undertake the evaluation of their innovations, even if they are inexperienced. The evaluator must, however, realize both his strengths and his limitations. This will enable him to work around his limitations and take full advantage of his strengths. In some cases, he may be able to reduce some of his limitations. If there is a college or university nearby, consultant help may possibly be obtained at little or no cost. In addition, courses in evaluation may be offered; and library resources may be available. If not, consultants can be obtained from university or professional research and evaluation

services. When hiring consultants, care should be taken to use their time wisely by supplying the consultant with pertinent literature before his visit and by determining how the consultant can be used to the best advantage. In either case, guidelines for the evaluation process, such as the one developed in this study, may be helpful in planning and carrying out the evaluation. Also, inquiries directed to other unified science projects may bring information concerning their evaluation experiences.

Other personnel limitations on the small-scale project include inadequate time, especially for data analysis, and inadequate support staff (clerical and secretarial help). A search of the local community may identify a number of people who are willing to donate their time for little or no remuneration. Arrangements might be made with a professor for students in an evaluation course to receive credit for assisting in an actual evaluation. A graduate student might be convinced to center his dissertation on a part of the evaluation. Retirees in many communities would enjoy the opportunity to do productive work. Another possible resource is the housewife who has had professional training, but cannot accept full-time employment responsibilities. She may be willing to help in evaluation activities, especially if she can work at home.

The constraints on small-scale, local unified science projects are numerous. But, there are also many ways in which the effects of these constraints can be lessened. Personal acquaintances, awareness of community resources, and ingenuity are important assets in attempting to overcome constraining factors. However, the major assets available to unified science projects are the innovative people involved who are

willing to give the extra effort and sacrifice necessary to make a project or an evaluation successful despite the apparent constraints.

In the process of optimizing the final purposes of the evaluation, the evaluator of the small-scale, local unified science project must consider the same factors that any other evaluator should consider: the potential purposes identified as being appropriate for the evaluation; the priorities placed on each of these purposes; the relative cost of each purpose in terms of time, effort, money, and effect of that purpose on other potential purposes; and the uncontrolled constraints of the situation. The relative cost of each purpose will depend primarily on the methods envisioned for gathering data concerning the purpose. But, the importance of these costs will depend upon the constraints of the situation. By determining how the uncontrolled constraints will affect each of the evaluative purposes, the evaluator can know how much evaluation he can afford in terms of the costs suggested above. Then, by weighing the cost of each purpose against the priority assigned to it, the evaluator can limit the evaluative purposes to those which are important to the project and to those which can yield appropriate information at low cost to the success of high priority activities. The determination of the optimum purposes is a somewhat involved task, but a very important one to assure that the evaluation does not attempt more than can be accomplished, and that the information obtained will be the most suitable for the decision questions pertinent to the evaluation.

The determination of the relationships among the purposes of the evaluation and between those purposes and the project itself is, also, likely to be complex. If the flow charting technique suggested in Chapter IV is applied to produce a model of the total development-

evaluation process, then it is suggested that the evaluators of multiyear, unified science projects make a general model of the entire process first. Then the process can be broken into phases that seem appropriate and examined in more detail without loosing sight of the overall plan.

In the process of design development, the evaluator must first consider the sources from which he will need to obtain information. Generally, the information sources available for a small-scale project are the same in nature but fewer in number than those for longer projects. Frequently, the evaluators of small-scale projects find that random assignment of students to experimental or control classes is not possible. More often, the evaluator must deal with classes that are self-selected or recruited into the experimental class. These factors can limit both the internal and external validity of the evaluative results, but the selection of appropriate evaluation designs and analysis techniques can help avoid these difficulties. The use of control classes that are similar, even if not randomly equivalent, to the experimental classes can improve the internal validity of the evaluation by controlling sources of error such as the effects of history, maturation, testing, instrumentation, regression, and experimental mortality on the evaluative results. The two groups can be statistically equated during analysis using analysis of covariance techniques. Even when the random assignment of students to control or experimental classes is not feasible, the randomization of testing may be. This means that one part of each class, control and experimental, is randomly chosen to take the pre-test and the other part to take the post-test. The advantage to this approach is an increase in the external validity (generalizability) of the results.

Sequential testing, in which data collection may occur at several times over the course of evaluation activities, is another possible approach to increasing internal validity. If possible, this testing might begin well before experimental treatment begins so that interaction effects of the treatment with other variables can be controlled.

The verification of initial results through repeated trials of the innovative curriculum materials with new students, especially in different settings, can also substantially improve the acceptability and generalizability of the evaluation. Finally, if an adequate description of the evaluation setting is provided, then interested school administrators will be able to determine for themselves how similar their own situations are to that of the evaluation and whether or not a change is worth the cost and risk involved.

The discussion of evaluation design has centered primarily on testing. In the evaluation of unified science programs, while tests are most useful in determining achievement in the cognitive and psychomotor domains, they are not as helpful in obtaining information concerning the affective domain. There are, however, many other data gathering techniques that can provide valuable information. Questionnaires can be administered using evaluation designs similar to those suggested for testing. They can be used to obtain information in any domain, but are of most use in the affective domains. Questionnaires may also be useful in obtaining information from or about the decision-makers and for the purposes of antecedent assessment. Observations can be very misleading if attempted on a "one shot" basis or even as a pre-test and post-test. Instead, multiple observations should be made so that consistent behavior

patterns can be identified. Otherwise, results may be confusing and inconclusive. In addition to the usefulness of observation in all of the domains of learning, this technique can provide information about the social, political, and physical environments as well. Interview techniques are mainly useful for obtaining affective information and information concerning the social setting and the needs of decision-makers. Projective techniques have been used for some time to gain insight into the psychological problems of individuals. They are now beginning to be used to obtain information concerning student attitudes and interests. These techniques seem promising for use in evaluating unified science goals concerning student attitudes and interests and concerning the understanding of interrelationships. Case study techniques at present do not seem to be appropriate for studying the effects of a unified science course on students. However, case study techniques are well suited for obtaining information for the assessment of developmental processes. Studies of this kind are needed to describe the "local" approach to curriculum development that has characterized unified science. Reviews are often approached on a "one shot" basis; however, this limits the possibility of change as new advances are made. The primary importance of review techniques to unified science project evaluations is in context assessment and in the formative and summative study of materials produced by the project.

According to Webb, Campbell, Swartz, and Sechrest, "Interviews and questionnaires must be supplemented by methods testing the same social science variables but having <u>different</u> methodological weaknesses " (Webb, Campbell, Swartz, and Sechrest, 1966, p. 1). The need for such "multiple operations" certainly applies to data-gathering techniques in

addition to interviews and questionnaires. If similar results are obtained from each of several such "multiple operations," then the validity of the evaluative information is strengthened significantly. The major supplementary data-gathering operation recommended by Webb, Campbell, Swarts, and Sechrest is the use of nonreactive, or unobtrusive, measures. These include any method for obtaining information in which there is little interaction between the evaluator and the original source of the information. The chance of bias produced by such interactions is minimized. Webb, Campbell, Swarts, and Sechrest identify three major types of unobtrusive measures; physical traces, archive records, and observations (Webb, Campbell, Swartz, and Sechrest, 1966). Some of the unobtrusive measures that might be of use in obtaining information about evaluation purposes for unified science include: class attendance records, records of the checkout of science books from the school library, student records, school newspapers, papers concerning science written for other classes, volunteered comments, observations of student involvement (noise level, clustering of students in laboratories, student initiation of studies or experiments), amount and type of wear on textbooks , and consumption of laboratory supplies. There are many more unobtrusive measures which evaluators will be able to identify from their own experiences. The exclusive reliance on unobtrusive measures for information regarding most evaluation purposes would be inappropriate and would result in incomplete data. Such measures can, however, provide strong supportive data, suggest explanations for the results obtained using other techniques, and identify phenomena that warrant additional examination.

While each of the data-gathering techniques can be used profitably in the evaluation of small-scale, local unified science programs, the availability of instruments suitable for such use is limited. This limitation is due primarily to the uniquenesses of unified science goals. The content goals of unified science programs are interdisciplinary. Consequently, standardized tests developed for individual disciplines are not entirely appropriate. Some of the other goals of Unified Science concern the development of a scientific attitude, interest in science, and skill in applying scientific processes to situations outside the classroom. These goals are at least as important and perhaps more important than content goals, but the achievement of these goals is less frequently evaluated. This lack of evaluation may be partly due to the difficulty in assessing the achievement of these goals and the less refined nature of the instruments that are used to obtain information of this type. It may also be partly due to the low saleability of instruments that are primarily of interest just to researchers and evaluators. With a few exceptions, commercial catalogues most frequently include achievement tests, aptitude tests, intelligence tests, and guidance tests of personality characteristics and vocational interest. Some of these will, no doubt, be of use to the unified science project evaluator. But, he will generally have to locate appropriate noncommercial instruments or develop instruments of his own.

The list below includes twenty-three instruments that either were developed for use in unified science evaluations, seem to hold promise for use with unified science programs, might be made applicable through revision, or appear to represent worthwhile guides to the development of new instruments.

1. "Rational Image of the Universe" (Slesnick, 1963, pp. 302-314). Developed for the evaluation of the Ohio State University Unified Science Program.

2. "Iowa Science and Culture Study Achievement Test" (Cossman and Fitch, 1967, pp. 683-686). Developed as part of the Iowa Science and Culture Study.

3. "Science Observation and Comparison Test" (Hungerford and Miles, 1969, pp. 61-66).

4. "Application of Principles Test-Science" (Progressive Education Association, 1939). Developed for the Eight-Year Study.

5. "Idea-Centered Laboratory Science Test" (Van Deventer, 1971). Developed for a junior high unified science program.

6. "Processes of Science Test" (Biological Sciences Curriculum Study Test Committee, 1965).

7. "Science Process Inventory" (Welch and Pella, 1967-1968, pp. 64-68).

8. "Empiricism Test" (Kaplan, 1963, pp. 341-350).

9. "Interest Index 8.2a" (Progressive Education Association, 1939). Developed for the Eight-Year Study.

10. "Nature of Science Scale" (Kimball, 1967-1968, pp. 110-120).

11. "What Is Science?" (Monona Grove High School Science Department, 1969). Developed for use with a unified science program, not as a research instrument.

12. "Test on Understanding Science" (Cooley and Klopfer, 1961). Available from the Educational Testing Service.

13. "Facts About Science Test" (Stice _et al_., 1958). Available from the Educational Testing Service.

14. "Science Opinion Survey" (Cossman and Fitch, 1967, pp. 683-686). Developed as part of the Iowa Science and Culture Study.

15. "Attitude Behavior Scale" (Progressive Education Association, 1939). Developed for the Eight-Year Study.

16. "Inventory of Scientific Attitudes" (Moore and Sutman, 1970, pp. 85-94).

17. "Battle Student Attitude Scale" (Battle, 1954).

18. "Vitrogen Science Attitude Scale" (Vitrogen, 1967, pp. 170-175).

19. "Harvard Project Physics Semantic Differential" (Geis, 1969). Developed for the evaluation of Harvard Project Physics.

20. "Portland Project Semantic Differential" (Portland Project, 1970). Developed for the evaluation of a unified science project, the Portland Project.

21. "Picture Story Test" (Parker, 1964). A self-concept projective instrument.

22. "Teacher Practices Observation Record" (Brown, 1968).

23. "Florida Taxonomy of Cognitive Behavior" (Soar et al., 1968).

Additional listings of published instruments can be found in the Mental Measurements Yearbooks, in Tests in Print, and in the Mental Measurements Yearbook Monograph Series. Many unpublished attitude scales are included in Scales for Measurement of Attitudes by Shaw and Wright. Unpublished and up-to-date instruments designed for the evaluation of science programs can be identified by frequent reference to professional science journals, especially the Journal of Research in Science Teaching and Science Education. Other articles concerning newly developed instruments can be located through the Education Index and the ERIC Current Index to Journals in Education. Test publisher's catalogues can be another source of information about current evaluative materials. In addition, queries directed to the Carl Campbell Brigham Library of Educational Testing Service, Princeton; to the University Microfilms Library Services; or to the ERIC Information Analysis Center for Science and Mathematics Education, Columbus, Ohio, may prove worthwhile if the evaluator can be fairly specific in his requests.

When considering an evaluative instrument for use in a specific evaluation, it is advisable that the evaluator obtain either a specimen set or a sample of the instrument well before its scheduled use. The titles of evaluative instruments can frequently be misleading. Only

through close examination can an evaluator be certain of the appropriateness of an evaluative instrument for the purposes of his evaluation.

The evaluator's alternative to purchasing commercial instruments or borrowing instruments constructed for other research or evaluation activities is constructing his own evaluative instruments. The reliability of such instruments will generally be open to question. But the use of multiple evaluating operations and the increased appropriateness of the instrument's use will often counterbalance the negative aspects of project-constructed instruments. Unobtrusive measures would be especially useful in this situation in providing supplementary data for purposes investigated with project-developed instruments.

The nature of the schedule developed to implement the evaluative design will depend upon the needs and situation of the particular unified science project and the schedule of the school in which the evaluation takes place. This first step in scheduling involves only an approximate determination of the dates on which the major evaluative activities, such as testing, will occur.

In small-scale, local projects the amount of logistical activity necessary for the smooth operation of the project will be reduced due to the proximity of the participants. However, logistical activities and other evaluative activities must still be scheduled carefully in order to insure that they will actually be carried out. This second step in scheduling must consider the evaluation plans in much greater detail. The program evaluation and review technique (PERT) is highly recommended for this purpose.

Even though both the participants and the evaluators of small-scale projects may occupy the same buildings or rooms, the maintenance of rapport is still essential. It may, in fact, be even more important. The fact that participating teachers are nearby and are seen frequently does not assure communication. Specific plans should be made to facilitate communication.

For the collection of data, the evaluative conditions should be as close to normal conditions as possible. If comparisons are to be made between experimental and control groups or among several different experimental treatments, then the evaluative environment at the time of data collection should be as similar as possible for all groups. Precautions that might be taken include the use of identical instructions for each group and the collection of data from each group on the same day and at the same time, if possible.

Analysis procedures for the evaluation of small-scale, local unified science programs differ in two respects from those of other projects. First, the information needs of decision-makers are related to the unique goals and characteristics of unified science programs as explained earlier in this chapter. Second, the nature of the evaluative sample, frequently small in number and nonrandom, will require the use of special analytical procedures (in addition to the special considerations made in the evaluative design) in order to obtain results that will be generalizable beyond the evaluative sample. Nonparametric techniques can be used to analyze data obtained from such samples. But, the evaluator should be aware that some nonparametric techniques are apt to be less sensitive to significant changes or significant differences between groups than are comparable parametric techniques.

In interpreting conclusions for the use of local decision-makers, the evaluator must keep in mind that the decision-maker may not be able to understand highly technical reports. Consequently, the evaluator must make his interpretations clear and precise. Yet, he must also avoid making generalizations beyond those which his data will support. The extent of generalizability of conclusions is determined by the evaluative situation and the design and instruments used in data collection. In addition, anticipation of possible outcomes of the evaluation in advance of interpretation of the conclusions will reduce interpretive biases.

The local nature of the project also affects the report, in that local decision-makers will probably not require highly formal reports. However, even if they are informal, the reports should be as complete as is useful and desired by the decision-maker. If the evaluator has identified informal decision-makers, then he must assure that these decision-makers also receive appropriate reports. Since decision-making in the evaluation of a small-scale, local unified science project will be mostly of a local nature, the evaluator will generally have ample opportunity to accompany his reports in presenting the evaluative information to decision-makers. In local decisions, the effects of internal and external pressures on decision-makers may be substantial. The evaluator's awareness of these pressures will facilitate the timely presentation of the report to reduce the effects of the pressures.

Summary

In summary, the process of curriculum development evaluation can be applied to small-scale, local unified science projects. These projects are unique in the impetuses that brought about their development, the goals of the developmental projects, and the characteristics of the

resulting unified science programs. The unique impetuses, goals, characteristics, and constraints of small-scale, local unified science projects significantly shape evaluation attempts. The evaluative possibilities are both exciting and challenging. The constraints on time, money, and personnel available to project evaluations are serious. If not fully considered, and eliminated whenever possible, these constraints could have a detrimental effect on the worth of the evaluation. However, there are many ways of reducing their effects and taking advantage of the unique opportunities available to the evaluator of a small-scale, local unified science project.

Additional sources of information concerning selected topics for further study are listed at the end of this chapter.

Keyed Additional References for Curriculum Development Evaluation

Key:

EP - Evaluation Process (includes many of the other topics listed)
RP - Research Process
PE - Purposes for Evaluation
IDM - Identification of Decision-Makers
SV - Societal Values
C - Constraints
PP - Project Plans
DGT - Data Gathering Techniques
ED - Evaluation Designs
SE - Sources of Error
DGI - Data Gathering Instruments
IS - Instrument Selection
IC - Instrument Construction
RBT - Rapport Building Techniques
ST - Scheduling Techniques
AM - Analysis Methods
S - Statistics
R - Reporting
DM - Decision-Making

References:

EP DGI — Beatty, Walcott H. Improving Educational Assessment and an Inventory of Measures of Affective Behavior. Washington, D. C.: Association for Supervision and Curriculum Development, NEA, 1969.

PE IC DGT — Bloom, Benjamin S.; Hastings, J. Thomas; and Madaus, George F. Handbook on Formative and Summative Evaluation of Student Learning. New York: McGraw-Hill Book Company, 1971.

DGT — Boyd, Robert D., and De Vault, M. V. "The Observation and Recording of Behavior," Review of Educational Research. XXXVI (No. 5) (December, 1966), Pp. 529-551.

DGI — Buros, Oscar K., ed. Tests in Print. Highland Park, New Jersey: Gryphon Press, 1961.

DGI — ________. The Sixth Mental Measurement Yearbook. Highland Park, New Jersey: Gryphon Press, 1965.

DM IDM SV — Cahill, Robert S., and Hencley, Stephen P., eds. The Politics of Education in the Local Community. Danville, Ill.: The Interstate Printers and Publishers, Inc., 1964.

ED SE — Campbell, Donald T., and Stanley, Julian C. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally and Company, 1963.

ST — Cook, Desmond L. Program Evaluation and Review Technique, Application to Education. U. S. Department of Health, Education and Welfare, OE-12024 Cooperative Research Monograph No. 17. Washington, D. C.: U. S. Government Printing Office, 1966.

S — Downie, N. M., and Heath, R. W. Basic Statistical Methods. 2nd. ed. New York: Harper and Row, 1965.

ST — Dresdner, David M.; Spiech, John A.; and Uslan, Gerald M. Federal Electric Corporation. A Programmed Introduction to PERT. New York: John Wiley and Sons, Inc., 1963.

RP AM R, DGT — Fox, David J. The Research Process in Education. New York: Holt, Rinehart and Winston, Inc., 1969.

DGT IC, SE, SV, ED, AM — Gage, N. E., ed. Handbook of Research on Teaching. Chicago: Rand McNally and Company, 1963.

DGT RBT — Garrett, Annette. Interviewing: Its Principles and Methods. New York: Family Service Association of America, 1942.

S — Glass, Gene V., and Stanley, Julian C. Statistical Methods in Education and Psychology. Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1970.

RP DGT R — Good, Carter V. Essentials of Educational Research. New York: Appleton-Century-Crofts, 1966.

DM — Gordon, Ira J. "A Model for Curriculum Development Decision-Making," ASCD News Exchange, May, 1967, Supplementary Edition; and in Gordon, Ira J. ed. Criteria of Theories of Instruction. Appendix B, pp. 40-44. Washington, D. C.: Association for Supervision and Curriculum Development, NEA, 1968.

EP — Grobman, Hulda. Evaluation Activities of Curriculum Projects: A Starting Point. (American Educational Research Association Monograph Series on Curriculum Evaluation, No. 2) Chicago: Rand McNally and Company, 1968.

PP C PE — ________. Developmental Curriculum Projects: Decision Points and Processes. Itasca, Ill.: F. E. Peacock Publishers, Inc., 1970.

S — Guilford, J. P. Fundamental Statistics in Psychology and Education. 4th ed. New York: McGraw-Hill Book Company, 1965.

AM — Hallenbeck, George P. "Some Uses of Test Data in Formative Evaluation," in Grobman, Hulda, Evaluation Activities of Curriculum Projects: A Starting Point. Appendix E, pp. 119-121. Chicago: Rand McNally and Company, 1968.

RP SE S — Hill, Joseph E., and Kerber, August. Models, Methods and Analytical Procedures in Educational Research. Detroit: Wayne State University Press, 1967.

DM IDM — Iannaccone, Laurence. Politics in Education. New York: The Center for Applied Research in Education, Inc., 1967.

DM IDM — Kimbrough, Ralph B. Political Power and Educational Decision-Making. Chicago: Rand McNally and Company, 1964.

DGT IS — Klopfer, Leopold E. "Evaluation of Learning in Science," in Bloom, Benjamin S. et al. Handbook on Formative and Summative Evaluation of Student Learning. "Chapter 18" pp. 561-641. New York: McGraw-Hill Book Company, 1971. (Table of Specifications for Science Education pp. 562-653).

PE — Lindvall, C. M., and Cox, Richard C. Evaluation as a Tool in Curriculum Development: The IPI Evaluation Program. (American Educational Research Association Monograph Series on Curriculum Evaluation, No. 5) Chicago: Rand McNally and Company, 1970.

DGT — Osgood, Charles E., Suci, George J., and Tannenbaum, Percy H. The Measurement of Meaning. Urbana, Ill. University of Illinois Press, 1966.

RP DGT, SE ED — Sax, Gilbert. Empirical Foundations of Educational Research. Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1968.

DGI — Shaw, Marvin E., and Wright, Jack M. Scales for Measurement of Attitude. New York: McGraw Hill Book Company, 1967.

IC AM SE — Thorndike, Robert L., and Hagan, Elizabeth. Measurement and Evaluation in Psychology and Education. New York: John Wiley and Sons, Inc., 1961.

EP DGT Tyler, Ralph W.; Gagne, Robert M.; and Scriven, Michael. Perspectives of Curriculum Evaluation. (American Educational Research Association Monograph Series on Curriculum Evaluation, No. 1) Chicago: Rand McNally and Company, 1967.

DGT SE Webb, E. J.; Campbell, Donald T.; Swartz, R. D.; and Sechrest, Lee. Unobtrusive Measures: Nonreactive Research in the Social Studies. Chicago: Rand McNally, 1966.

CHAPTER VI

A SYSTEMS MODEL FOR THE EVALUATION PROCESS

The problems encountered by the evaluators of curriculum development projects indicate the need for ways to organize information concerning evaluation theory into frameworks of manageable size and testable nature. Also indicated is the need for integrating constructs to tie these frameworks together so that the evaluation of curriculum development projects can be viewed as an operational system. One way to organize and integrate material of a theoretical nature is through the use of general systems theory.

General systems theory is based on the belief that it is possible to represent all forms of inanimate matter as systems (Hass, 1967, p. 1). English and English define "system" as "...the set of orderly and persisting interrelations between parts of a whole" (English and English, 1958, p. 541). According to K. E. Boulding, the objectives of general systems theory are:

> ...to point out the similarities in the theoretical constructions of different disciplines, where these exist, ...to develop theoretical models having applicability to at least two different fields of study...and to develop something like a "spectrum" of theories--a system of systems which may perform the function of a "gestalt" in theoretical construction (Boulding, 1956, p. 2).

Systems can be differentiated in two ways: by the level of abstraction used as a model (pictorial, descriptive, abstract-mathematical), or by the metaphor employed (machine, organism, field). Systems also

can be identified as either "closed" or "open." The closed system is isolated from the environment and, consequently, obeys the law of increasing entropy; i.e., the system will continually increase its disorganization. A laboratory experiment in chemistry would be an example of a closed system in which the number of variables is kept low by isolation. A system that is related to and exchanges matter with the environment is called an "open" system (e.g., living organisms). Thus, such a system must have input and output of energy and information. With these inputs and outputs and a method of self-regulation, the system can balance the entropy increase and maintain a steady state. This self-regulation, called feedback, distinguishes the system as an organismic, open system.

Every system (except the smallest) is composed of subsystems, and, conversely, all subsystems but the largest are parts of a suprasystem (the system in its environment). Each subsystem contains the alternatives necessary to delineate all aspects of the concept it represents. Also, each subsystem can be characterized by particular functions it performs and by its relation to each of the other subsystems. Thus, there is a dynamic interplay of subsystems operating as functional processes. Every system has a somewhat arbitrary boundary which distinguishes the system from the environment, the environment being everything external to the system. The proximal environment is that part of the environment of which the system is aware. The distal environment is the part of the environment that affects the operation of a system, but is beyond the awareness of the system. Finally, there are factors in the system and the environment which will affect the structure and the function of the system. If these factors are located

inside the system, they are known as variables. If they are located in the environment, they represent the effects that the environment can have on the system and are called parameters (Hearn, 1960).

The thesis of this chapter is that general systems theory can provide the basis for the formation of an adequate model for the process of curriculum development evaluation. In developing this model there are three sets of criteria to which special attention must be paid. First, from the model, we should be able to produce an evaluation which is composed of certain elements common to all curriculum development evaluations, whatever their design. These elements include the ones identified in Chapter III:

I. Context Identification
   a. Identification of Criteria
   b. Identification of Constraints
II. Optimization
III. Design Development
IV. Implementation
V. Analysis
VI. Reporting
VII. Decision-Making

Second, it must be theoretical in nature (i.e., consisting of a group of related, comprehensive, general characterizations). Thus, the model must meet the criteria of comprehensiveness (i.e., characterizing the conditions of the situation), and generality (i.e., applicable to any time and place)(Maccia and Maccia, 1966, p. 117). Third, the model must be capable of producing a "practice theory." A "practice theory" is generic to a particular set of problems, which is grounded in particular scientific concepts, and which has as its purpose

the improvement of professional practice. This practice theory has three elements: (1) a scientific component based on scientific theory which in turn must be based on empirically collected data and must be logical (i.e., shows the relationships between facts in a logical manner); (2) an artistic component; and (3) an intuitive component.

Figure 14 illustrates the components of a general systems model. The system consists of an environment, a control apparatus (the general system), a number of control elements (subsystems), an input, an output, and a feedback network. The integrating construct used is the process of curriculum development evaluation. The developing curriculum may be one designed for use in a school or school system. However, to include in the system the evaluation of programs designed to be used in more than a single school system will lead to numerous difficulties. The problems confronted become increasingly vague and complicated as the size increases. Consequently, the resulting evaluation plan would then have to become more general and less specific.

The "universe of discourse" represents the total environment associated with the system. The "general system" represents that portion of the environment which has direct bearing upon the formation of a theory. Contained in that part of the "universe" not included in the "general system" are those factors which have an indirect bearing on the theory, providing the "parameters" that limit the extent of the general system.

Another part of the environment outside of the "general system" is the sources from which "available data" are drawn. The term "available data " refers to the set of information obtained from the proximal environment through research or some other form of investigation. "Available data" are taken into the "general systems" by a

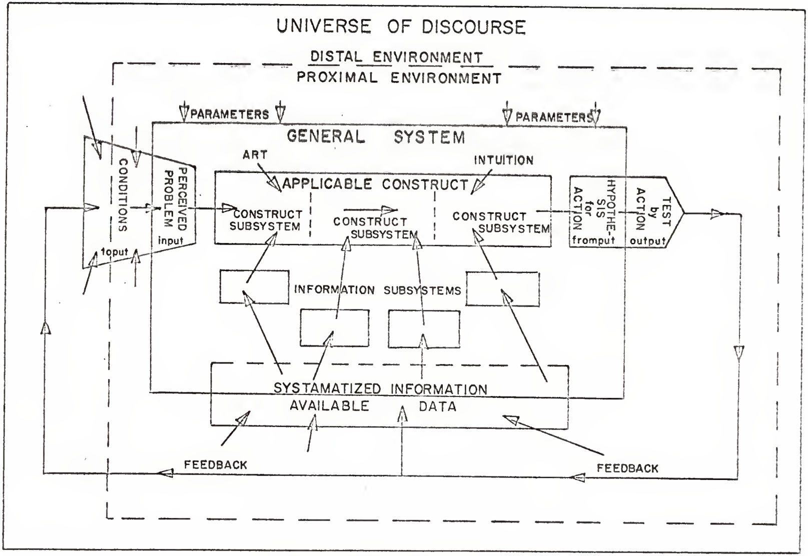


Figure 14.--General Systems Model

process of "systematizing" and become "systematized information." The information is then sorted into various "information subsystems." Depending on the nature of the data, they may, in fact, become part of any number of subsystems. This process forms the scientific basis of the system. Data are empirically collected and then logically related. Since data can often fit into more than one subsystem, there is an extensive overlap and interrelation between the subsystems. Each time additional data enter a subsystem, a part of that system is either verified or modified. Thus, the system is dynamic.

A third part of the environment located outside of the "general system" is responsible for the "toput conditions." The "toput conditions" are all of the factors in the environment that define a particular problem which confronts the system. The "input" is the "perceived problem" as viewed by those persons who are functioning as part of the system. Not all of the "toput" can become "input" (a part of the system), because the perceptual limitations of the persons operating the system do not permit them to recognize all of the applicable conditions. In other words, part of these conditions are located in the distal environment.

Thus, there are three different influences of the environment on the system. First, are the "conditions" which are applicable only to a particular situation confronting the system. Second, are the "parameters." These are applicable either to a certain locale for an extended period, or to a large area for a relatively short period of time. Factors that function as parameters move in and out of the system. Third, are the "available data" applicable to any situation at any time.

The "applicable construct" consists of a series of "construct subsystems." Through logical thought, art, intuition, and the use of various information subsystems, the optimum solution to the perceived problem is identified and applied. The result is "fromput," information produced by the system that is ready to enter the environment. Such "fromput" can be described as an "hypothesis for action." "Fromput" enters the environment as action (the "output") and is tested by the environment. The results of the testing form feedback to the "available data," eventually entering the system and altering or verifying the information subsystems. Feedback also affects the "toput conditions," eventually changing the perception of problems by the persons operating in the system. The feedback process, by providing a test for hypotheses, is then a fourth way in which the environment acts upon the system. The self-regulation of this system by continually improving the state of its own information and by altering its own conditions identifies it as an "organismic open system."

Figure 15 illustrates the application of general systems theory to the process of curriculum development evaluation. The "toput" for an evaluation includes the criteria and constraints for that evaluation. The evaluator's perception of the criteria and the constraints results in the selection of the purposes of the evaluation and a model of the evaluation plan. These purposes and plans become "input" into the system in a process that parallels Optimization. The applicable construct which ties the systems together is the central part of the process of curriculum development evaluation, consisting of three construct subsystems: Design Development, Implementation, and Analysis. As the process progresses, various bits of information are required by the

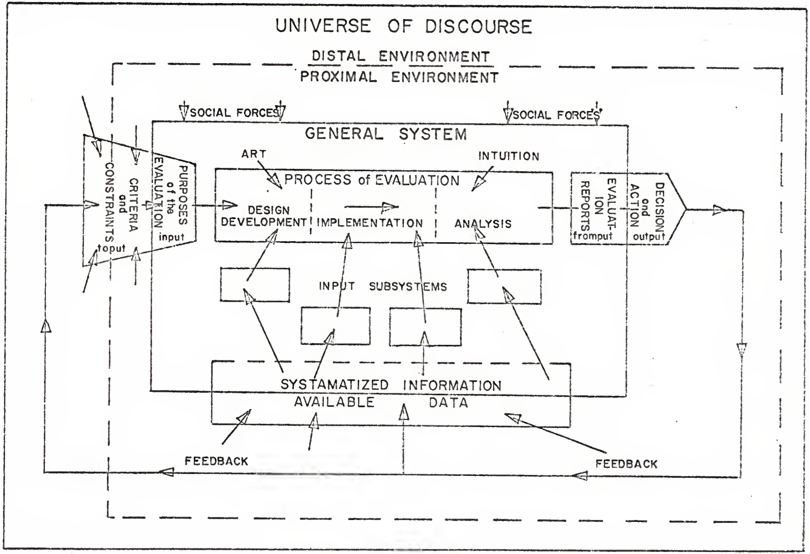


Figure 15.--Systems Model for Curriculum Development Evaluation

evaluator concerning the needs, the alternatives, and the influences which are related to his project evaluation. The evaluator is free to select the most applicable combinations of information from the appropriate "input subsystem." These "input subsystems" are identical with the Input into each of the component processes discussed in Chapter IV. The use of logical thought, art, and intuition aid the evaluator in making optimum selections as well as in interpreting the information collected in the system. Evaluation reports are the "fromput" or hypotheses for action from the evaluation process. As these reports move from the evaluation system into the environment, they are acted upon by decision-makers in the environment. The decision-makers determine the actions that will be taken in the environment relative to the project evaluated. Part of the actions will form feedback which modified the evaluation system by producing new needs, alternatives, influences, criteria, and constraints. The environmental parameters of the evaluation system include contemporary social forces such as the problems of the knowledge explosion, changing values, and social unrest. The proximal environment consists of the collection of decision-makers and the decision needs, alternatives, and influences of which the evaluator is aware. In the distal environment are the decision-makers that are not apparent to the evaluator. As the evaluator becomes familiar with the environment in which an evaluation occurs, the boundaries of the proximal environment will be pushed outward so that the evaluator is more aware of the decision-making process that affects the curriculum development project being evaluated. Consequently, the evaluation will become more responsive to its environment.

Summary

The preceding chapter has considered the evaluation process on a theoretical level through the use of general systems theory. The flow chart approach developed in Chapters III and IV viewed the evaluation process on a more practical level. A general systems model provides an indication of what should be, while the flow chart model indicates what is or is not. Use of a general systems model of evaluation provides insight into how the evaluation process should function as changes needed to improve the process become evident. Flow charting illustrates how an evaluation proceeds and provides evidence of needed changes in evaluative practices. Thus, the systems model presented in this chapter will be of most use to the person concerned with evaluation theory, while the flow chart model will be of most use to the practitioner. Each, however, aids in understanding the evaluation process and, hopefully, will lead to the improvement of evaluation practice.

CHAPTER VII

SUMMARY AND IMPLICATIONS

This study has determined the need for guidelines to aid in the evaluation of small-scale, local unified science curriculum development projects. The development of guidelines is now feasible because of recent advances in theory concerning the process of curriculum development evaluation. A flow chart model was presented to provide evaluators with an overall view of the evaluation process. Investigation of the details of the model led to insight into the operation of the various components of the evaluation process in general and specifically in relation to the unique criteria and constraints of the small-scale, local unified science project. Finally, a general systems model was used to examine the process of evaluation and its relation to its total environment on a theoretical level.

From this study a number of conclusions and implications can be drawn concerning the evaluation of small-scale, local unified science projects, the nature of the evaluation process, the use of the model in the improvement of the evaluation process, research on various aspects of the evaluation process, and the educational preparation of curriculum evaluators.

1. The use of this model can provide guidelines for the evaluation of small-scale, local unified science programs.

This study could encourage evaluation activities connected with small-scale, local unified science projects and could aid in making

decisions among alternatives through the consideration of the needs and influences of the situation, thus improving both the quantity and the quality of these evaluations.

Adaptation of the model for use in other types of small-scale projects may encourage their evaluation and provide guidelines for the accomplishment of the evaluation.

Adaptation of the model could be useful to schools considering the adoption of new materials through aiding them in conducting their own evaluation of the appropriateness of those materials.

2. Local unified science projects have unique criteria and serious constraints.

There is a need for clear recognition of constraints so that they can be (1) circumvented through the identification of mitigating factors; (2) considered as realistic limitations to the evaluation and appropriate adjustments made in the evaluative purposes; or (3) acknowledged, but without adjustments made, so that their effects on the evaluation process can be studied.

There is a serious need for additional evaluative instruments for use in the evaluation of unified science programs.

All alternatives should be considered in making decisions concerning evaluative purposes and methods.

No single evaluation plan can be developed that is applicable to all unified science projects.

3. Evaluation processes vary little from project to project. However, input into these processes will vary greatly.

Evaluation plans will vary from project to project.

Because process is relatively more stable than input or output, more emphasis should be placed on the improvement of evaluative process.

There is a need to improve methods of determining unique input and identifying alternative input when possible. However, no one set of universal input can be established.

4. There are many decisions to be made during the process of a curriculum development project; they will be confronted at various stages in the project.

Evaluation should be an integral part of the curriculum development process rather than imposed upon it. Evaluation activities should begin with the beginning of the project.

5. The primary intent of evaluation is to aid in decision-making, not to test hypotheses, which is a major intent of research.

There is a need for consideration of how evaluation designs should differ from research designs.

There is a new role for the evaluator in providing appropriate information for decision-making. He must get to know the decision-maker and communicate with him frequently.

There is a need for the development of methods for determining the "practical significance" of evaluative information to be used by the various decision-makers.

There is a need for flexibility in evaluation so that it can be responsive to changes in needs or influences.

6. Complex evaluational processes can be represented with flow chart diagrams for consideration on the practical level, and with general systems models for the study of evaluation on the theoretical level.

Consideration of the use of these techniques to describe processes in other areas of science education might be profitable.

There is a need for further investigation into the representation of the process of evaluation using flow charting and general systems techniques.

7. Present efforts in evaluation theory tend to concentrate on limited portions of the evaluation process (e.g., on the use of one or two evaluative purposes or on particular types of instrumentation). However, a clear understanding of the total process requires an overall view as well.

There is a need for additional theoretical consideration of the entire process of evaluation.

8. The use of flow charting techniques can provide a framework for the identification of areas of present weakness in the various components of the evaluation process on a practical level.

There is a need for extensive study of the optimization process, both in clarifying methods for determining optimum processes and in developing models of the curriculum project development evaluation process.

There is a need for additional information about methods of reporting other than formal research reports, and on how to select the method of reporting appropriate for each decision-maker.

There is a need for the study of the increased use of graphical techniques in data analysis.

Additional information is needed concerning the use of nonparametric techniques for the analysis of data drawn from "natural" settings.

The practical aspects of evaluation implementation (rapport building, logistics, scheduling, budgeting, and data organization) need to be considered more fully.

There is a need for more guidance in the construction of instruments, especially for techniques other than testing.

There is a need for new approaches to commercial instrument production. Commercial instruments are often inappropriate for an evaluator's needs. This implies the need for more flexible approaches, such as commercial test-item pools.

9. The examination of the evaluation process through the use of flow charting techniques and general systems models can lead to the identification of areas rich in questions that need to be considered in the research of these various aspects of the process of evaluation.

There is a critical need for research on the decision-making process in curriculum development projects. Research questions that seem fruitful include: What sources of information other than the evaluation report are used by each of the various decision-makers? What pressures are operating on the decision-makers, and with what effects? How do various decision-makers view the evaluation process and how do their views affect their decisions? If decision rules are established, to what extent are they followed? Are the decision-makers identified the ones who make the decisions; if not, who really makes the decisions? How effective are the actions initiated by the various decision-makers? Can any type of decision or decision-maker be identified as being of higher priority or more significant than others, or does this vary with the project or the personality of the person who is the decision-maker?

Research concerning the process of reporting is needed to determine the relationship between the type of report, the decision-maker involved, and the effect of the report on the decision.

Research is needed concerning the relative costs of various evaluative activities. This refers to cost in terms of money, time, effort, and the effect on other evaluative activities.

Research is needed on the effects of various constraining factors on the evaluation activities and on the evaluative results.

Research is needed on the effects of evaluation on the biases on the evaluative activities and results.

10. There is presently little provision in the education of curriculum developers for the study of the process of evaluation.

This implies a need for the development of evaluation courses (in addition to research courses) to prepare and encourage curriculum developers to undertake worthwhile evaluations of their efforts.

The study also implies that the process of curriculum development evaluation is an integral part of the process of curriculum development and that the education of evaluators should not be separated from the education of curriculum developers or teachers.

BIBLIOGRAPHY

Alcorta, Louis B. "Relative Effectiveness of Two Organizational Patterns for Science Instruction in Secondary Schools," Dissertation Abstracts. Vol. 23 Stanford University (1962), pp. 2030-2031.

Alkin, Marvin C. "Evaluation Theory Development," Evaluation Comment. Vol. 2 (No. 1) (June, 1970), pp. 2-7.

Battle, J. A. "Techniques and Instruments for Measuring Certain Student Human Relations." Ed.D. Dissertation, College of Education, University of Florida, 1954.

Biological Sciences Curriculum Study Test Committee. "Processes of Science Test." New York: The Psychological Corporation, 1965.

Brown, Bob Burton. The Experimental Mind in Education. New York: Harper and Row, 1968.

Brown, Edward K. "More Effective Educational Research Through the Use of Process Evaluation Techniques," Journal of Research and Development in Education. Vol. 3 (No. 4) (Summer, 1970), pp. 59-71.

Boulding, Kenneth E. "General Systems Theory: The Skeleton of Science," from Management Science. Abington, Pa.: Institute of Management Science, April, 1956. (Mimeographed.)

Campbell, Donald T., and Stanley, Julian C. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally and Company, 1963.

Cooley, W. W., and Klopfer, L. E. "Test on Understanding Science (TOUS)." Princeton, N. J.: Educational Testing Service, 1961.

Cossman, George W. "The Effects of a Course in Science and Culture for Secondary School Student," Journal of Research in Science Teaching. Vol. 6 (No. 3) (1969), pp. 274-283.

________., and Fitch, Robert M. "The Iowa Science and Culture Study," Social Education. Vol. 31 (Dec., 1967), pp. 683-686.

DeRose, James V. "Theory into Action - Contributions to a Unified Science Program," The Science Teacher. Vol. 32 (May, 1965), p. 83.

Dresdner, David M.; Spiech, John A.; and Uslan, Gerald M. A Programmed Introduction to PERT. Prepared for the Federal Electric Corporation. New York: John Wiley and Sons, Inc., 1963.

English, Horace B., and English, Ava Champney. A Comprehensive Dictionary of Psychological and Psychoanalytical Terms. New York: David McKay Company, Inc., 1958.

Federation for Unified Science Education. "Unified Science Education Projects." Monona Grove, Wisconsin, Sept. 26, 1969a. (Mimeographed.)

Federation for Unified Science Education. Minutes of "What can FUSE do to best serve schools presently involved with unified science programs." Chaired by Michael Fiasca; Fourth Annual FUSE Conference, Monona Grove, Wisconsin, Sept. 27, 1969b.

Fiasca, Michael A. "Feasibility of Integration of Selected Aspects of (CBA) Chemistry, (CHEMS) Chemistry, and (PSSC) Physics into a Two Year Physical Science Sequence," Dissertation Abstracts. Vol. 27 (No. 8) Oregon State University (1967), p. 2349-A.

Forehand, Garlie A. "Curriculum Evaluation as Decision-Making Process," Journal of Research and Development in Education. Vol. 3 (No. 4) (Summer, 1970), pp. 27-37.

Fontaine, Thomas D. "Federal Programs for the Improvement of School Science and Mathematics," Science Education. Vol. 54 (Sept., 1970), pp. 209-212.

Fox, David J. The Research Process in Education. New York: Holt, Rinehart and Winston, Inc., 1969.

Fox, Fred W. "Education and the Spirit of Science--The New Challenge," The Science Teacher. Vol. 33 (Nov., 1966), pp. 58-59.

Gadsden, Thomas, Jr. "A Quasi-Experimental Study of the Effectiveness of the First Year of a Three Year Integrated Science Program Offered to Tenth Grade Students at P. K. Yonge Laboratory School." Paper prepared for Educational Research, EDF-760 at the University of Florida, Gainesville, Florida, June, 1968. (Typewritten.)

Geis, Fred, Jr. "The Semantic Differential Technique as a Means of Evaluating Changes in 'Affect'." Ph.D. Dissertation, Harvard University, Boston, Massachusetts, 1969.

Grobman, Hulda. Evaluation Activities of Curriculum Projects: A Starting Point. American Educational Research Association Monograph Series on Curriculum Evaluation No. 2. Chicago: Rand McNally and Company, 1968.

________. Developmental Curriculum Projects: Decision Points and Processes. Itasca, Illinois: F. E. Peacock Publisher's Inc., 1970.

Guba, Egon C. "The Failure of Educational Evaluation," Educational Technology. Vol. 9 (No. 5) (May, 1969), pp. 29-38.

Hand, Harold C. "Children and Youth--Pawns or Sacred Entities?" 1966, pp. 1-15. (Periodical name and date not available.)

Hass, C. Glenn. "General Systems Theory," Gainesville: University of Florida, 1967. (Mimeographed.)

Hearn, G. Theory Building in Social Work. Toronto: University of Toronto Press, 1960.

Hungerford, Harold R., and Miles, David T. "A Test to Measure Observation and Comparison Skills in Science," Science Education. Vol. 53 (No. 1) (Feb., 1969), pp. 61-66.

Hurd, Paul De Hart. "New Directions in Science Teaching," The Education Digest. Vol. 32 (March, 1967), pp. 16-19.

Jacobson, Willard J. "Approaches to Science Education Research: Analysis and Criticism," Journal of Research in Science Teaching. Vol. 7 (No. 3) (Fall, 1970), pp. 217-225.

Kaplan, E. H. "Empiricism as a Function of Training in the Scientific Method," Journal of Research in Science Teaching. Vol. 1 (1963), pp. 341-350.

Kimball, Merritt E. "Understanding the Nature of Science: A Comparison of Scientists and Science Teachers," Journal of Research in Science Teaching. Vol. 5 (1967-68), pp. 110-120.

Klopfer, Leopold E. "Integrated Science for the Secondary School: Process, Progress, and Prospects," The Science Teacher. Vol. 33 (Nov., 1966), pp. 27-31.

________., and McCann, Donald C. "Evaluation in Unified Science: Measuring the Effectiveness of the Natural Science Course at the University of Chicago High School," Science Education. Vol. 53 (No. 2) (March 1969), pp. 155-164.

Lerner, Morris R. "Integrated Science," The Science Teacher. Vol. 31 (Feb., 1964), pp. 37-38.

Light, Richard J., and Smith, Paul V. "Choosing a Future: Strategies for Designing and Evaluating New Programs," Harvard Educational Review. Vol. 40 (No. 1) (Feb., 1970), pp. 1-28.

Lindvall, C. M., and Cox, Richard C. Evaluation as a Tool in Curriculum Development: The IPI Evaluation Program. American Education Research Association Monograph Series on Curriculum Evaluation No. 5. Chicago: Rand McNally and Company, 1970.

Maccia, Elizabeth S., and Maccia, George S. Development of Educational Theory Derived from Three Educational Theory Models. Columbus, Ohio: The Ohio State University Research Foundation, Dec., 1966.

Monona Grove High School Science Department. "What Is Science?" Monona, Wisconsin: 1969. (Mimeographed.)

McCloskey, James. "The Development of the Role of Science in General Education for Elementary and Secondary Schools," Dissertation Abstracts. Vol. 24 University of North Carolina, Chapel Hill (1963), pp. 3194-3195.

Moore, Richard W., and Sutman, Frank X. "The Development, Field Test and Validation of an Inventory of Scientific Attitudes," Journal of Research in Science Teaching. Vol. 7 (No. 2) (1970), pp. 85-94.

Neurath, Otto et al. "Encyclopedia and Unified Science," International Encyclopedia of Unified Science. Vol. 1 (No. 1). Chicago: University of Chicago Press, 1938.

Parker, James. "The Relation of Self Report to Self-Report to Inferred Self Concept in Sixth Grade Children." Ed.D. Dissertation, College of Education, University of Florida, 1964.

Pfeiffer, Carl. "Evaluation of Unified Science," Federation for Unified Science Education Bulletin. No. 12. August, 1970, p. 4.

Portland Project. "Portland Project Semantic Differential." Portland, Oregon: 1970. (Mimeographed.)

Progressive Education Association. "Evaluation in the Eight-Year Study." Chicago: University of Chicago, 1939. (Mimeographed.)

Randall, Robert S. "An Operational Application of the CIPP Model for Evaluation," Educational Technology. Vol. 9 (No. 7) (July, 1969), pp. 40-44.

Sax, Gilbert. Empirical Foundations of Educational Research. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1968.

Scriven, Michael. "The Methodology of Evaluation," Perspectives of Curriculum Evaluation. Edited by Robert Stake. American Educational Research Association Monograph Series on Curriculum Evaluation No. 5. Chicago: Rand McNally and Company, 1967.

Shamos, Morris H., and Barnard, J. D. "Evaluation of COPES Activities," COPES Newsletter. (No. 4) March, 1970.

Showalter, Victor M. "A Summary of Effects of a Unified Science Curriculum on High School Graduates." Paper summarizing the final report for U. S. Office of Education Grant OEG-1-7-068937-3761, 1968. (Mimeographed.)

________. "The Promise of Unified Science Education," Science Education News. March, 1969, pp. 1-2.

________., and Thompson, Barbara, eds. Federation for Unified Science Education Bulletin. No. 12. August, 1970.

Slesnick, Irwin L. "The Effectiveness of a Unified Science in the High School Curriculum," Journal of Research in Science Teaching. Vol. 1 (No. 4) (1963), pp. 302-314.

Smith, Herburt A., and Washton, Nathon S. "Science in the Secondary School," Review of Educational Research. Vol. 27 (Oct., 1957), pp. 343-355.

Soar, Robert et al. "Florida Taxonomy of Cognitive Behavior." Gainesville: University of Florida, 1968. (Mimeographed.)

Stake, Robert E. "Language Rationality, and Assessment." Improving Educational Assessment and an Inventory of Measurement of Affective Behavior. Edited by W. H. Beaty. Washington, D. C.: Association for Supervision and Curriculum Development, NEA, 1969, pp. 14-40.

Stice, Glen et al. "The Facts About Science Test." Princeton, New Jersey: Educational Testing Service, 1958.

Stufflebeam, Daniel L. "Evaluation as Enlightenment for Decision Making," Improving Educational Assessment and an Inventory of Measures of Affective Behavior. Edited by W. H. Beaty. Washington, D. C.: Association for Supervision and Curriculum Development, NEA, 1969, pp. 41-73.

Van Deventer, W. C. "A New Look at Inquiry in Junior High School Science." Kalamazoo, Michigan: Western Michigan University, Jan., 1971. (Mimeographed.)

Vitrogan, David. "A Method for Determining a Generalized Attitude of High School Students Toward Science," Science Education. Vol. 51 (March, 1967), pp. 170-175.

Webb, E. J.; Campbell, Donald T.; Swartz, R. D.; and Sechrest, Lee. Unobtrusive Measures: Nonreactive Research in the Social Studies. Chicago: Rand McNally, 1966.

Welch, Wayne W., and Pella, Milton O. "The Development of an Instrument for Inventorying Knowledge of the Process of Science," Journal of Research in Science Teaching. Vol. 5 (1967-1968), pp. 64-68.

Welch, Wayne W., and Walbert, Herbert J. "A Design for Curriculum Evaluation," Science Education. Vol. 52 (No. 1) (Feb., 1968), pp. 10-16.

Westbury, Ian. "Curriculum Evaluation," Review of Educational Research. Vol. 40 (No. 2) (April 1970), pp. 239-260.

BIOGRAPHICAL SKETCH

Thomas Gadsden, Jr., was born May 19, 1944, at Thomasville, Georgia. In 1962, he was graduated from Mainland High School, Daytona Beach, Florida. In April, 1966, he received the degree Bachelor of Science with a major in Physics, and in August, 1967, the degree Master of Education, with a major in Science Education and a minor in Physics, both from the University of Florida.

From September, 1967, until September, 1969, he taught half-time at P. K. Yonge Laboratory School, University of Florida, and from September, 1969, until the present, full-time, while pursuing his work toward the degree Doctor of Education.

Thomas Gadsden, Jr., is married to the former Sandra Victoria Howard, and is the father of a daughter, Julianne Christin. He is a member of the American Association of Physics Teachers, Federation for Unified Science Education, National Science Teachers Association, Association for Supervision and Curriculum Development, National Association for Research in Science Teaching, the John Dewey Society, Phi Delta Kappa, and Phi Kappa Phi.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Education.

L. A. Arnold

Luther A. Arnold, Chairman
Associate Professor of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Education.

Vynce A. Hines

Vynce A. Hines
Professor of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Education.

Richard E. Garrett

Richard E. Garrett
Professor of Physics

This dissertation was submitted to the Dean of the College of Education and to the Graduate Council, and was accepted as partial fulfillment of the requirements for the degree of Doctor of Education.

August, 1971

B. J. Shaf by M C Beller

Dean, College of Education

Dean, Graduate School

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Education.

Daniel C. Swanson
Professor Emaritus of Physics